

Search completed: May 6, 2004, 09:30:46
Job time : 52.7244 secs

```
XX WO200252015-A2.
PN
XX
XX 04-JUL-2002.
PD
XX
XX 20-DEC-2001; 2001WO-CA001843.
PF
XX
XX 22-DEC-2000; 2000US-0257857P.
PR
XX
XX (BOEHR ) BOEHRINGER INGELHEIM CANADA LTD.
PA
XX
XX Kukulj G, Pause A;
PI
XX
XX MPI; 2002-575382/61.
DR
XX N-PSDB; ABX8587.
XX
XX
XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.
XX
XX Disclosure; Page 120-129; 140pp; English.
XX
XX The invention describes a self-replicating hepatitis C virus (HCV)
XX CC polynucleotide molecule comprising a 5'-non translated region (NTR),
XX CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
XX CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-
XX CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
XX CC potential inhibitors of HCV replication. The HCV RNA molecule is also
XX CC useful for efficiently establishing cell culture replication. The self-
XX CC replicating polynucleotide molecule contains a 5'-NTR, where G at
XX CC position 1 is substituted for A, and therefore provides an alternative to
XX CC existing systems comprising a self-replicating HCV RNA molecule that, in
XX CC conjunction with mutations in the HCV non-structural region, such as the
XX CC G(2042)C/R mutations, transduces and/or replicates with greater
XX CC efficiency. This amino acid sequence and/or replicates with the hepatitis C virus
XX CC replicon ApG12 and contains the viral protease NS2/3, protease complex
XX CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
XX Sequence 2201 AA;
SQ
Query Match 95.0%; Score 1887; DB 5; Length 2201;
Best Local Similarity 93.4%; Pred. No. 9.9e-174;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
QY 2 LTTSPYKYLARLIMWLOVLTREAHQWIPPLNVRGGRDAIILLTCAVHPBELLFD 61
DB 19 LTTSPHYKLFARLIMWLOVLTREAHQWIPPLNVRGGRDAVILLTCAIHPBELLFT 78
QY 62 ITKLLAIFGFLWVLAQGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
DB 79 ITKLLAIFGFLWVLAQGITKVPYFVRAHGLIRACMLVRKAGGHYVQMALMKLAALTGT 138
QY 122 YVYDHLTPPLQWMAAGRLDAVAEPIFSDMEVKIITWGADTAAGDIIISGLPVSARG 181
DB 139 YVYDHLTPPLQWMAAGRLDAVAEPIFSDMEVKIITWGADTAAGDIIISGLPVSARG 198
QY 182 REIILGPADNFEQGWRLAPITAYSQQTRGLIGCIITSLTGRDKNQVEGEVQVYSTAQ 241
DB 199 REIILGPADNFEQGWRLAPITAYSQQTRGLIGCIITSLTGRDRNQVEGEVQVYSTAQ 258
QY 242 SFLATCVGVCWTYFEGAGSKTLAAGPKPIQWNTYNTDODLVGQAPRGASMTPECTCGS 301
DB 259 SFLATCVGVCWTYFEGAGSKTLAAGPKPIQWNTYNTDODLVGQAPRGASMTPECTCGS 318
QY 302 SDLYVTRHADVIPIRRRGDSRGSILSPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 361
DB 319 SDLYVTRHADVIPIRRRGDSRGSILSPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 378
QY 362 GVAKAADPIPYESMETIMR 380
DB 379 GVAKAADPIPYESMETIMR 397
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OW protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 12.3758 seconds
(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-12
Perfect score: 1997
Sequence: 1 ALLTSPYKVLARLIMWL.....RGVAKAVDFIPVESMETTR 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1902	95.7	3010	1 GNMVCI	genome polypotein
2	1890	95.1	3010	1 A45573	genome polypotein
3	1870	94.1	3010	1 GNMVTV	genome polypotein
4	1845	92.9	3010	1 S18030	genome polypotein
5	1823	91.7	3010	1 GNMVTC	genome polypotein
6	1715	86.3	3011	1 GNMVVC	genome polypotein
7	1709	86.0	3011	1 S40770	genome polypotein
8	1703	85.7	3011	1 GNMVCH	genome polypotein
9	1482	74.6	3014	1 UT5620	genome polypotein
10	1391	70.0	3033	1 GQ1303	genome polypotein
11	1391	70.0	3033	1 GNMVVC	genome polypotein
12	408.5	20.6	3005	2 T08841	polypotein - marm
13	342.5	17.2	2970	2 T08839	polypotein - marm
14	112	5.6	692	2 H71426	hypothetical prote
15	102.5	5.2	660	2 VHMW42	structural protein
16	101	5.1	564	2 S36637	signal recognition
17	101	5.1	600	2 B46642	DNA-directed DNA p
18	100.5	5.1	353	2 B46732	conserved hypotet
19	99	5.0	399	2 A43038	conserved hypotet
20	99	5.0	399	2 C98247	hypothetical 50.8K
21	97.5	4.9	1085	2 T03531	coBN protein homol
22	95.5	4.8	2796	2 JG4098	tetracycline 6-hyd
23	94.5	4.8	2796	2 JG4743	fatty-acid synthas
24	93.5	4.7	1380	2 A41509	conserved hypotet
25	93.5	4.7	1380	2 T18309	receptor-adenylate
26	93	4.7	7463	2 T36248	CD4 peptide synch
27	92.5	4.7	706	2 S33761	transferrin precu
28	92.5	4.7	716	2 G83612	hypothetical prote
29	92.5	4.7	3069	2 H70656	fatty-acid synthas

30.	92	4.6	659	1 B44212	structural protein
31	91.5	4.6	428	2 AF0241	probable coenzyme
32	91.5	4.6	665	2 D83252	nucleotide sugar e
33	91.5	4.6	3414	1 GNMVNE	genome polypotein
34	91	4.6	434	2 G82728	conserved hypotet
35	91	4.6	904	2 A84212	hypothetical prote
36	90.5	4.6	868	2 H81775	aconitate hydratase
37	90.5	4.6	871	2 H72597	hypothetical prote
38	89.5	4.5	961	2 AE0375	probable cation-tr
39	88.5	4.5	401	1 A36961	conserved hypotet
40	88.5	4.5	446	2 AE1150	probable nitrate r
41	88.5	4.5	652	2 C70688	hypothetical prote
42	88.5	4.5	2121	2 T27406	fatty acid synthas
43	88.5	4.5	3076	2 A87058	genome polypotein
44	87.5	4.4	3412	1 GNMVTV	genome polypotein
45	87	4.4	418	2 H90679	probable transport

ALIGNMENTS

RESULT 1
GNMVCI genome polypotein - hepatitis C virus (strain J)
N/contains: capsid protein C; envelope protein E; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C/Accession: A39253; PS0086
R/Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A/Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A/Reference number: A39253; MUID:91088550; PMID:2175903
A/Accession: A39253
A/Molecule type: genomic RNA
A/Residues: 1-3010 <KAT>
A/Cross-References: GB1D90208; NID:9221610; P1DN:BA14233.1; P1D:9221611
R/Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A/Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A/Reference number: PS0085
A/Accession: PS0086
A/Molecule type: genomic RNA
A/Residues: 2650-2707 <KAT>
A/Experimental source: Japanese isolate
C/Comment: The cleavage sites of this polypotein have not been determined.
C/Superfamily: hepatitis C virus genome polypotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypotein; serine
F/2-115/Product: capsid protein C #status predicted <CP>
F/116-191/Product: envelope protein M #status predicted
F/192-389/Product: major envelope protein E #status predicted <ME>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1112-1317/Region: nucleotide-binding motif B
F/1116-1319/Region: DEKH motif
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match	95.7%	Score 1902;	DB 1;	Length 3010;
Best Local Similarity	94.2%	Pred. No. 6.9e-150;		
Matches 357;	Conservative 12;	Mismatches 10;	Indels 0;	Gaps 0;
OY	2	LLTSPYKVLARLIMWLQYITRVAHLQVWIPPLNVRGGDAIILLTCAVHPLIFD	61	
DB	828	LLTSPYKVLARLIMWLQYITRVAHLQVWIPPLNVRGGDAIILLTCAVHPLIFD	887	
OY	62	ITKLLALRPLNVLQAGIKVYFRAOGLIACVLVKAAGHVVQVAFMGLALTGT	121	
DB	888	ITKLLALRPLNVLQAGIKVYFRAOGLIACVLVKAAGHVVQVAFMGLALTGT	947	

QY 122 YVYDHLTPLODMWAGRLDVAVEPIVFSMEVXITTWGADTAACGDIISGLPVSARRG 181
DB 948 YVYDHLTPLODMWAGRLDVAVEPIVFSMEVXITTWGADTAACGDIISGLPVSARRG 1007
QY 182 REILLGADNFEQGMRLAPITAYSOQTRGLGCIITSLTGRDKNVGESEVQVVSATQ 241
DB 1008 KEILLGADNFEQGMRLAPITAYSOQTRGLGCIITSLTGRDKNVGESEVQVVSATQ 1067
QY 242 SFATCNGVCWTVFHAGSKTLAGPKGPIITQYNTVDDLVGMQAPPGARSMTPTCTCGS 301
DB 1068 SFATCNGVCWTVFHAGSKTLAGPKGPIITQYNTVDDLVGMQAPPGARSMTPTCTCGS 1127
QY 302 SDLYLVRHADVPVRRGDSRGLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 361
DB 1128 SDLYLVRHADVPVRRGDSRGLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 1187
QY 362 GYAKAVDPIPVESMETTR 380
DB 1188 GYAKAVDPIPVESMETTR 1206

RESULT 2

A:Accession: A45573
M: genome polypeptide - hepatitis C virus (strain J7)
N: Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
C: Species: hepatitis C virus
C: Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C: Accession: A45573
R: Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A: Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: E
A: Reference number: A45573; MUID:9225714; PMID:118627
A: Accession: A45573
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-3010 <TRAN>
A: Cross-references: GB:D11168; GB:D01171; NID:9221612; PIDN:BA01943.1; PID:9221613
A: Experimental source: HCV-JT
A: Note: Sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
C: Superfamily: hepatitis C virus genome polypeptide
C: Keywords: ATP; glycoprotein; hydroxylase; nucleotide binding; P-loop; polypeptide; serin
F: 116-191/Product: capsid protein C #status predicted <CPC>
F: 116-191/Product: envelope protein M #status predicted <EPW>
F: 192-389/Product: major envelope protein E #status predicted <MEB>
F: 390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F: 730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F: 1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F: 1230-1237/Region: nucleotide-binding motif A (P-loop)
F: 1312-1317/Region: nucleotide-binding motif B
F: 1316-1319/Region: DEXH motif
F: 1616-1862/Product: nonstructural protein NS4 #status predicted <N4B>
F: 1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F: 2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match

Best Local Similarity

95.1%; Score 1890; DB 1; Length 3010;

Matches 357; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKYLARLIMWLOYLITRVEAHQWIPPLNVRGGDAIILLTCAVHEPLIFD 61
DB 828 LTLSPYKYLARLIMWLOYLITRVEAHQWIPPLNVRGGDAIILLTCAVHEPLIFD 887
QY 62 ITKLLAIFGLPMLVQAGITKVPYFRAAGLIRACMLVKAAGSHVQMAFMKLAALTGT 121
DB 888 ITKLLAIFGLPMLVQAGITKVPYFRAAGLIRACMLVKAAGSHVQMAFMKLAALTGT 947
QY 122 YVYDHLTPLODMWAGRLDVAVEPIVFSMEVXITTWGADTAACGDIISGLPVSARRG 181
DB 948 YVYDHLTPLODMWAGRLDVAVEPIVFSMEVXITTWGADTAACGDIISGLPVSARRG 1007
QY 182 REILLGADNFEQGMRLAPITAYSOQTRGLGCIITSLTGRDKNVGESEVQVVSATQ 241

DB 1008 REILLGADNFEQGMRLAPITAYSOQTRGLGCIITSLTGRDKNVGESEVQVVSATQ 1067
QY 242 SFATCNGVCWTVFHAGSKTLAGPKGPIITQYNTVDDLVGMQAPPGARSMTPTCTCGS 301
DB 1068 SFATCNGVCWTVFHAGSKTLAGPKGPIITQYNTVDDLVGMQAPPGARSMTPTCTCGS 1127
QY 302 SDLYLVRHADVPVRRGDSRGLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 361
DB 1128 SDLYLVRHADVPVRRGDSRGLSPRPVSYLKSGSGGFLCPSGHAVGIFRAAVCTR 1187
QY 362 GYAKAVDPIPVESMETTR 380
DB 1188 GYAKAVDPIPVESMETTR 1206

RESULT 3

A:Accession: A40244
M: genome polypeptide - hepatitis C virus (strain Taiwan)
N: Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
C: Species: hepatitis C virus
C: Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C: Accession: A40244
R: Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A: Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A: Reference number: A40244; MUID:92230206; PMID:11314449
A: Accession: A40244
A: Molecule type: genomic RNA
A: Residues: 1-3010 <CHE>
A: Cross-references: GB:M64754
C: Superfamily: hepatitis C virus genome polypeptide
C: Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydroxylase; nonstructural
F: 116-191/Product: capsid protein C #status predicted <CPC>
F: 116-191/Product: envelope protein M #status predicted <EPW>
F: 192-389/Product: major envelope protein E #status predicted <MEB>
F: 390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F: 730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F: 1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F: 1230-1237/Region: nucleotide-binding motif A (P-loop)
F: 1312-1317/Region: nucleotide-binding motif B
F: 1316-1319/Region: DEXH motif
F: 1616-1862/Product: nonstructural protein NS4 #status predicted <N4B>
F: 1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F: 2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F: 196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match

Best Local Similarity

94.1%; Score 1870; DB 1; Length 3010;

Matches 349; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKYLARLIMWLOYLITRVEAHQWIPPLNVRGGDAIILLTCAVHEPLIFD 61
DB 828 LTLSPYKYLARLIMWLOYLITRVEAHQWIPPLNVRGGDAIILLTCAVHEPLIFD 887
QY 62 ITKLLAIFGLPMLVQAGITKVPYFRAAGLIRACMLVKAAGSHVQMAFMKLAALTGT 121
DB 888 ITKLLAIFGLPMLVQAGITKVPYFRAAGLIRACMLVKAAGSHVQMAFMKLAALTGT 947
QY 122 YVYDHLTPLODMWAGRLDVAVEPIVFSMEVXITTWGADTAACGDIISGLPVSARRG 181
DB 948 YVYDHLTPLODMWAGRLDVAVEPIVFSMEVXITTWGADTAACGDIISGLPVSARRG 1007
QY 182 REILLGADNFEQGMRLAPITAYSOQTRGLGCIITSLTGRDKNVGESEVQVVSATQ 241
DB 1008 REILLGADNFEQGMRLAPITAYSOQTRGLGCIITSLTGRDKNVGESEVQVVSATQ 1067
QY 242 SFATCNGVCWTVFHAGSKTLAGPKGPIITQYNTVDDLVGMQAPPGARSMTPTCTCGS 301
DB 1068 SFATCNGVCWTVFHAGSKTLAGPKGPIITQYNTVDDLVGMQAPPGARSMTPTCTCGS 1127

QY 302 SDLYLVRHADVI PVRRRDSRGSLLSPRPVSLKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 1128 SDLYLVRHADVI PVRRRDSRGSLLSPRPVSLKSSGGPILCPGSHAVGIFRAAVCTR 1187
QY 362 GVAKAVDFIPVESMETTR 360
DB 1188 GVAKAVDFIPVESMETTR 1206

RESULT 4

SI8030
genome polyprotein - hepatitis C virus (isolate JKI)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Variety: isolate JKI
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text_change 23-Mar-2001
C:Accession: SI8030; S33570; A48332; SI8029
R:Honda, M.; Kaneko, S.; Nasashin, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A:Reference number: SI8028
A:Accession: SI8030
A:Molecule type: genomic RNA
A:Residues: 1-3010 <HON>
A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479
A:Experimental source: isolate JKI from an individual
R:Honda, M.; Kaneko, S.; Uenura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A:Reference number: A48332; MUID:93119270; PMID:8380322
A:Accession: S33570
A:Molecule type: genomic RNA
A:Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'D', 655-761, 'T', 763-782 <HON>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:P.121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EMP>
F/192-389/Product: major envelope protein E #status predicted <MES>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1307-615/Product: hepacivirin #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif A (P-loop)
F/1316-1319/Region: DEXH motif
F/1316-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 92.9%; Score 1845; DB 1; Length 3010;
Best Local Similarity 91.8%; Pred. No. 3.9e-145;
Matches 349; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 1 ALTLSPYKYLARLIMWLQYLITRVEAHLQVIMPELNVGRGDAIILITCAVHPELIF 60
DB 827 ALTLSPYKYLARLIMWLQYLITRVEAHLQVIMPELNVGRGDAIILITCAVHPELIF 886
QY 61 DITLLAIIFGLPVLVLAQGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 120
DB 887 DITLLAIIFGLPVLVLAQGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 946
QY 121 TYVVDHLTPLODMAHAGLRDLAAVEVTFSDMEVKIITWGADTPAACGDIISGLPVARR 180
DB 947 TYVVDHLTPLODMAHAGLRDLAAVEVTFSDMEVKIITWGADTPAACGDIISGLPVARR 1006
QY 181 GREILLGPADNFBEGQWRLAPITAYSQOTRGILGCIITSLTGRDKQVGEQVAVSTAT 240

DB 1007 GREILLGPADNFBEGQWRLAPITAYSQOTRGILGCIITSLTGRDKQVGEQVAVSTAT 1066
QY 241 QSFATCVNGCWTVEFHGAGSKTLAGRKGPITQWNTYTVDDLVGMQAPFGARSTPTCTG 300
DB 1067 QSFATCVNGCWTVEFHGAGSKTLAGRKGPITQWNTYTVDDLVGMQAPFGARSTPTCTG 1126
QY 301 SSDLYLVRHADVI PVRRRDSRGSLLSPRPVSLKSSGGPILCPGSHAVGIFRAAVCTR 360
DB 1127 SSDLYLVRHADVI PVRRRDSRGSLLSPRPVSLKSSGGPILCPGSHAVGIFRAAVCTR 1186
QY 361 GVAKAVDFIPVESMETTR 360
DB 1187 GVAKAVDFIPVESMETTR 1206

RESULT 5

GNMVTIC
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J
J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from human
A:Reference number: A38465; MUID:91140698; PMID:1847440
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TKX>
A:Cross-references: EMBL:M58335; NID:G329770; PIDN:AA12945.1; PID:G329771
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: major envelope protein E #status predicted <EMP>
F/192-389/Product: major envelope protein E #status predicted <MES>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: hepacivirin #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1316-1862/Product: nonstructural protein NS4a #status predicted <N4a>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 91.7%; Score 1823; DB 1; Length 3010;
Best Local Similarity 91.0%; Pred. No. 2.7e-143;
Matches 345; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 2 LTLSPYKYLARLIMWLQYLITRVEAHLQVIMPELNVGRGDAIILITCAVHPELIF 61
DB 828 LTLSPYKYLARLIMWLQYLITRVEAHLQVIMPELNVGRGDAIILITCAVHPELIF 887
QY 62 ITKLLAIIFGLPVLVLAQGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 121
DB 888 ITKLLAIIFGLPVLVLAQGITKVPYFVRAQGLIRACMLVKAAGHYVQMAFMKLAALTG 947
QY 122 YVVDHLTPLODMAHAGLRDLAAVEVTFSDMEVKIITWGADTPAACGDIISGLPVARR 181
DB 1008 YVVDHLTPLODMAHAGLRDLAAVEVTFSDMEVKIITWGADTPAACGDIISGLPVARR 1067
QY 182 RETLLGPADNFBEGQWRLAPITAYSQOTRGILGCIITSLTGRDKQVGEQVAVSTAT 241
DB 1068 RETLLGPADNFBEGQWRLAPITAYSQOTRGILGCIITSLTGRDKQVGEQVAVSTAT 1127
QY 242 SFLATCVNGCWTVEFHGAGSKTLAGRKGPITQWNTYTVDDLVGMQAPFGARSTPTCTG 301
DB 1068 SFLATCVNGCWTVEFHGAGSKTLAGRKGPITQWNTYTVDDLVGMQAPFGARSTPTCTG 1127
QY 302 SDLYLVRHADVI PVRRRDSRGSLLSPRPVSLKSSGGPILCPGSHAVGIFRAAVCTR 361

Db 1128 SDLVTRHADVTPVRRRDSSGSLSPRPVSYLKSSGGPILCFPGHNVGIFRAVCTR 1187
 QY 362 GVAQAVDFIPVESMETTMR 380
 Db 1188 GVAQAVDFIPVESMETTMR 1206

RESULT 6

GNWVC3

genome polypeptide - hepatitis C virus (strain HCV-1)
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C/Species: hepatitis C virus
 C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
 C/Accession: A39166; PID:91172826; PMID:1848704
 R/Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
 A/Title: Genetic organization and diversity of the hepatitis C virus.
 A/Reference number: A39166; MUID:91172826; PMID:1848704
 A/Accession: A39166
 A/Molecule type: genomic RNA
 A/Residues: 1-3011 <CHO>
 A/Cross-references: GB:M62221; NID:G229873; PIDN:AAA5676.1; PID:G229874
 R/Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I
 J. Gen. Virol. 73, 1131-1141, 1992
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A/Reference number: P00393; MUID:92268871; PMID:1316939
 A/Accession: P00403
 A/Molecule type: genomic RNA
 A/Residues: 1577-1633 <CHA>
 A/Cross-references: DDBJ:D10128
 A/Experimental source: isolates E-b16
 A/Accession: P00404
 A/Status: preliminary
 A/Molecule type: genomic RNA
 A/Residues: 1577-1633 <CH2>
 A/Experimental source: isolates E-b17

C/Superfamily: hepatitis C virus genome polypeptide
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 86.3%; Score 1715; DB 1; Length 3011;
 Best Local Similarity 82.8%; Pred. No. 2.8e-134;
 Matches 313; Conservative 32; Mismatches 33; Indels 0; Gaps 0;

QY 3 LTLSPYKVLARIWLQYLITRVEAHQVWIPPLNVGGRDAIILTCVAHPELIFDI 62
 Db 829 LTLSPYKVLARIWLQYLITRVEAHQVWIPPLNVGGRDAIILTCVAHPELIFDI 888
 QY 63 TKLLAIFGPMVLQGLITKVPYFVARAOGILRACMLVRAGAHVQMAFMKLAALTRY 122
 Db 889 TKLLAIFGPMVLQGLITKVPYFVARAOGILRACMLVRAGAHVQMAFMKLAALTRY 948
 QY 123 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 182
 Db 949 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 1008
 QY 183 EILIGPADNFEQGMRLAPITAYAOQTRGLGCIITSLTGDKXNVEGEVQVISTAAQS 242
 Db 1009 EILIGPADNFEQGMRLAPITAYAOQTRGLGCIITSLTGDKXNVEGEVQVISTAAQS 1068

QY 243 FLATCNGVCTVFFHAGSKTLAGPKPITQMTYNTVDOLVGMQAPPGAREMTPTCTGSS 302
 Db 1069 FLATCNGVCTVFFHAGSKTLAGPKPITQMTYNTVDOLVGMQAPPGAREMTPTCTGSS 1128
 QY 303 DLVYTRRADVTPVRARRDSSGSLSPRPVSYLKSSGGPILCFPGHNVGIFRAVCTR 362
 Db 1129 DLVYTRRADVTPVRARRDSSGSLSPRPVSYLKSSGGPILCFPGHNVGIFRAVCTR 1188
 QY 363 GVAQAVDFIPVESMETTMR 380
 Db 1189 GVAQAVDFIPVESMETTMR 1206

RESULT 7

S40770

genome polypeptide - hepatitis C virus
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C/Species: hepatitis C virus
 C/Date: 19-May-2000 #sequence revision 19-May-2000 #text_change 19-Jan-2001
 C/Accession: S40770; P01285
 R/Okamoto, H.
 submitted to the EMBL Data Library, March 1992
 A/Reference number: S40770
 A/Accession: S40770
 A/Molecule type: genomic RNA
 A/Residues: 1-3011 <OKA>
 A/Cross-references: EMBL:D10749; NID:G221586; PIDN:BA01582.1; PID:G221587
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
 Jpn. J. Exp. Med. 60, 167-177, 1990
 A/Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A/Reference number: P01284; MUID:91013116; PMID:2170712
 A/Accession: P01285
 A/Molecule type: genomic RNA
 A/Residues: 1-513 <OK2>
 A/Cross-references: GB:D00831; NID:G221511; PIDN:BA00705.1; PID:G221512
 A/Experimental source: isolate HC-J1
 C/Superfamily: hepatitis C virus genome polypeptide
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 86.0%; Score 1709; DB 1; Length 3011;
 Best Local Similarity 82.3%; Pred. No. 8.8e-134;
 Matches 311; Conservative 32; Mismatches 35; Indels 0; Gaps 0;

QY 3 LTLSPYKVLARIWLQYLITRVEAHQVWIPPLNVGGRDAIILTCVAHPELIFDI 62
 Db 829 LTLSPYKVLARIWLQYLITRVEAHQVWIPPLNVGGRDAIILTCVAHPELIFDI 888
 QY 63 TKLLAIFGPMVLQGLITKVPYFVARAOGILRACMLVRAGAHVQMAFMKLAALTRY 122
 Db 889 TKLLAIFGPMVLQGLITKVPYFVARAOGILRACMLVRAGAHVQMAFMKLAALTRY 948
 QY 123 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 182
 Db 949 VYDHLTPLODMAHAGRLDAVAEPIFSDMEVKIITWGADTAACGDIISGLPVSARRGR 1008
 QY 183 EILIGPADNFEQGMRLAPITAYAOQTRGLGCIITSLTGDKXNVEGEVQVISTAAQS 242
 Db 1009 EILIGPADNFEQGMRLAPITAYAOQTRGLGCIITSLTGDKXNVEGEVQVISTAAQS 1068
 QY 243 FLATCNGVCTVFFHAGSKTLAGPKPITQMTYNTVDOLVGMQAPPGAREMTPTCTGSS 302

DB 1069 FLATCINGCMTVYHAGRTTASPKGPIQMTNVNDQDVGMPAPGQSRSLTPTCTGSS 1128
 QY 303 DLYVTRHADVPYRRGDSRGSLLSPRVSYLKSSGGGLPCPSGHAVGIFRAAVCTRG 362
 DB 1129 DLYVTRHADVPYRRGDSRGSLLSPRVSYLKSSGGGLPCPSGHAVGIFRAAVCTRG 1188
 QY 363 VAKAVDFIPVESMETTR 380
 DB 1189 VAKAVDFIPVESMETTR 1206

RESULT 8

GNMVCH
 genome polypeptide - hepatitis C virus (strain H)
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural protein NS4; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A/Note: host Homo sapiens (man)
 C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C/Accession: A36814; A41546
 R:Inchuspe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Submitted to Genbank, July 1992
 A/Description: Genomic structure of the human prototype strain H of hepatitis C virus:
 A/Reference number: A36814
 A/Accession: A36814
 A/Molecule type: genomic RNA
 A/Residues: 1-3011 <INC>
 A/Cross-references: GB:M67463; NID:G329737; PIDN:AAA4534.1; PID:G329738
 R:Inchuspe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A/Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison
 A/Reference number: A41546; MUID:92052256; PMID:1658800
 A/Contents: annotation
 A/Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polypeptide; glycoprotein; hydrolyase; nonstructural
 C/Keywords: ATP; capsid protein C; status predicted <CPC>
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EMP>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis virus status predicted <NS3>
 F:1230-1237/Product: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 85.7%; Score 1703; DB 1; Length 3011;
 Best Local Similarity 82.3%; Pred. No. 2.8e-113;
 Matches 311; Conservative 34; Mismatches 33; Indels 0; Gaps 0;
 QY 3 LTLSPYVYVLLARLIMVLYITRVEALQVWIPPLNVRGGRDAIILLTCVHPELFDI 62
 DB 829 LTLSPYVYVLLARLIMVLYITRVEALQVWIPPLNVRGGRDAIILLTCVHPELFDI 888
 QY 63 TKLLAIFGPIMLVLOAGITKPYFVRAGLIRACLVYKKAAGHYVQMAFMKALATGT 122
 DB 889 TKLLAIFGPIMLVLOAGITKPYFVRAGLIRACLVYKKAAGHYVQMAFMKALATGT 948
 QY 123 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRG 182
 DB 949 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRG 1008
 QY 183 ELLGPADEFGQGRRLAPITAYSQOGRGLIGCTITSLTGDKQVGEVGVSTANOS 242
 DB 1009 ELLGPADEFGQGRRLAPITAYSQOGRGLIGCTITSLTGDKQVGEVGVSTANOS 1068
 QY 243 FLATCINGCMTVYHAGRTTASPKGPIQMTNVNDQDVGMPAPGQSRSLTPTCTGSS 302

DB 1069 FLATCINGCMTVYHAGRTTASPKGPIQMTNVNDQDVGMPAPGQSRSLTPTCTGSS 1128
 QY 303 DLYVTRHADVPYRRGDSRGSLLSPRVSYLKSSGGGLPCPSGHAVGIFRAAVCTRG 362
 DB 1129 DLYVTRHADVPYRRGDSRGSLLSPRVSYLKSSGGGLPCPSGHAVGIFRAAVCTRG 1188
 QY 363 VAKAVDFIPVESMETTR 380
 DB 1189 VAKAVDFIPVESMETTR 1206

RESULT 9

JCS620
 genome polypeptide - hepatitis C virus (isolate EUH1480)
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural protein NS4; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A/Note: the translation of the nucleotide sequence is not complete in this paper
 C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C/Accession: JCS620
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, J.A.; Simmonds, P.; Elliott, R.M.
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997
 A/Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
 A/Reference number: JCS620; MUID:97366593; PMID:9223423
 A/Accession: JCS620
 A/Molecule type: mRNA
 A/Residues: 1-3014 <CHN>
 A/Cross-references: GB:Y13184
 A/Experimental source: genotype 5a, which predominates in South Africa
 A/Note: the translation of the nucleotide sequence is not complete in this paper
 C:Superfamily: hepatitis C virus genome polypeptide
 C/Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polypeptide; serine
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EMP>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:384-408/Region: hypervariable #status predicted
 F:590-730/Product: nonstructural protein NS1 #status predicted <NS1>
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1008-1616/Product: hepatitis virus #status predicted <NS3>
 F:1231-1238/Region: nucleotide-binding motif A (P-loop)
 F:1313-1318/Region: nucleotide-binding motif B
 F:1317-1320/Region: DEXH motif
 F:1617-1863/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 74.6%; Score 1482; DB 1; Length 3014;
 Best Local Similarity 69.4%; Pred. No. 7.7e-115;
 Matches 263; Conservative 56; Mismatches 60; Indels 0; Gaps 0;
 QY 2 LTLSPYVYVLLARLIMVLYITRVEALQVWIPPLNVRGGRDAIILLTCVHPELFDI 61
 DB 829 LTLSPYVYVLLARLIMVLYITRVEALQVWIPPLNVRGGRDAIILLTCVHPELFDI 888
 QY 62 TKLLAIFGPIMLVLOAGITKPYFVRAGLIRACLVYKKAAGHYVQMAFMKALATGT 121
 DB 889 TKLLAIFGPIMLVLOAGITKPYFVRAGLIRACLVYKKAAGHYVQMAFMKALATGT 948
 QY 122 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRG 181
 DB 949 VYDHLTPLODMAHAGLRDLAAVAEVPVIFSDMEVKIITGADTAACGDIISGLPVASRRG 1008
 QY 182 ELLGPADEFGQGRRLAPITAYSQOGRGLIGCTITSLTGDKQVGEVGVSTANOS 241
 DB 1009 ELLGPADEFGQGRRLAPITAYSQOGRGLIGCTITSLTGDKQVGEVGVSTANOS 1068
 QY 242 FLATCINGCMTVYHAGRTTASPKGPIQMTNVNDQDVGMPAPGQSRSLTPTCTGSS 301
 DB 1069 FLATCINGCMTVYHAGRTTASPKGPIQMTNVNDQDVGMPAPGQSRSLTPTCTGSS 1128
 QY 302 SDLYVTRHADVPYRRGDSRGSLLSPRVSYLKSSGGGLPCPSGHAVGIFRAAVCTRG 361
 DB 1129 SDLYVTRHADVPYRRGDSRGSLLSPRVSYLKSSGGGLPCPSGHAVGIFRAAVCTRG 1188

QY 362 GVAKAVDFIPVESMETTKR 380
 DB 1189 GVAKALFVPEVNETTKR 1207

RESULT 10

Genome polyprotein - hepatitis C virus (isolate HC-J6)
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus genome having poor homology to report
 C/Species: hepatitis C virus
 C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
 C/Accession: J01303
 R/Okamoto, H.; Okada, S.; Sugiyama, Y.; Kura, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
 J. Gen. Virol. 72, 2697-2704, 1991
 A/Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
 A/Reference number: J01303; MUID:92044440; PMID:1658196
 A/Accession: J01303
 A/Molecule type: genomic RNA
 A/Residues: 1-303 <OKA>
 A/Cross-references: GB:D00944; NID:G221650; PIDN:BA00792.1; PID:G221651
 A/Experimental source: isolate HC-J6 from a Japanese individual
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein M #status predicted <EMP>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1316-1321/Region: nucleotide-binding motif B
 F:1320-1323/Region: DEXH motif
 F:1620-1866/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 70.0%; Score 139.1; DB 1; Length 3033;
 Best Local Similarity 66.0%; Pred. No. 3e-107;

Matches 250; Conservative 55; Mismatches 74; Indels 0; Gaps 0;

QY 2 LTLSPYKYLALRLWLYLITRVAHLQVWIPPLVNRGGDAIILITCAVHPELLD 61
 DB 832 LFTLTPYKTLRLRFWMWLCYLLTLEAWQVAPMVOVGRDGIIMAVAFCEGVVD 891
 QY 62 ITKLALIFGLPVLVQAGITKVPYFVRAQGLIRACMLYKKAAGHYVQMAFMKLAALTG 121
 DB 892 ITKMLAVIGPAVLKAGLITRVYFVRAHLLRMCTVMHLAGRYVQWVLLALGKWTGT 951
 QY 122 YVVDHITPLQDVAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARRG 181
 DB 952 YVVDHITPLQDVAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARRG 1011
 QY 182 RETLLGPADNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 241
 DB 1012 RETLLGPADNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 1071
 QY 242 SFPLATCVNVCWTFVHAGSKTLAPKPGITQMTNNVDOLVGMQAPGASMPCTGCG 301
 DB 1072 SFPLATCVNVCWTFVHAGSKTLAPKPGITQMTNNVDOLVGMQAPGASMPCTGCG 1131
 QY 302 SLDYLVTRADVTPVRRGDSRGLSPRPVSYLKGSSGGLPLCSGNAVGFRAAVCTR 361
 DB 1132 SLDYLVTRADVTPVRRGDSRGLSPRPVSYLKGSSGGLPLCSGNAVGFRAAVCTR 1191
 QY 362 GVAKAVDFIPVESMETTKR 380
 DB 1192 GVAKSIDIFVETLIVTR 1210

RESULT 11

GNVW38

Genome polyprotein - hepatitis C virus (strain HC-J6)
 N/Contains: capsid protein C; envelope protein M; hepatitis C virus genome having poor homology to report
 C/Species: hepatitis C virus
 C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C/Accession: A40250; P00397; P00559
 R/Okamoto, H.; Kura, K.; Okada, S.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.;
 Virol. 188, 331-341, 1992
 A/Title: Full-length sequence of a hepatitis C virus genome having poor homology to report
 A/Reference number: A40250; MUID:92230232; PMID:1314459
 A/Accession: A40250
 A/Molecule type: genomic RNA
 A/Residues: 1-3033 <OKA>
 A/Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BA01761.1; PID:G221609
 R/Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.
 J. Gen. Virol. 73, 1131-1141, 1992
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to ex
 A/Reference number: P00397; MUID:92268871; PMID:1316939
 A/Accession: P00397
 A/Molecule type: genomic RNA
 A/Residues: 2678-2754 <CHN>
 A/Cross-references: DDBJ:D10134
 A/Experimental source: isolate B-b12
 R/Kato, N.; Ootsuyama, Y.; Okoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991
 A/Title: Distribution of plural HCV types in Japan.
 A/Reference number: P00554; MUID:92068204; PMID:1720309
 A/Accession: P00554
 A/Molecule type: mRNA
 A/Residues: 2678-2729 <KAT>
 A/Cross-references: GB:D10542; GB:D90518; NID:G221523; PIDN:BA01418.1; PID:G221524
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein M #status predicted <EMP>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1234-1241/Region: nucleotide-binding motif A (P-loop)
 F:1316-1321/Region: nucleotide-binding motif B
 F:1320-1323/Region: DEXH motif
 F:1620-1866/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23;

Query Match 70.0%; Score 139.1; DB 1; Length 3033;
 Best Local Similarity 64.2%; Pred. No. 3e-107;

Matches 244; Conservative 64; Mismatches 72; Indels 0; Gaps 0;

QY 1 ALTLSPYKYLALRLWLYLITRVAHLQVWIPPLVNRGGDAIILITCAVHPELLD 60
 DB 831 SIFLTPYKTLRLSRVWMLSTYVLAEOIQWVPLEVRGGRDGIIMAVAFCEGVVD 890
 QY 61 DITKLALIFGLPVLVQAGITKVPYFVRAQGLIRACMLYKKAAGHYVQMAFMKLAALTG 120
 DB 891 DITKLALIFGLPVLVQAGITKVPYFVRAQGLIRACMLYKKAAGHYVQMAFMKLAALTG 1011
 QY 121 TYVVDHITPLQDVAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARR 180
 DB 951 TYVVDHITPLQDVAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARR 1010
 QY 181 GRELLGPADNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 240
 DB 1011 GRELLGPADNFEQGGRLAPITAYSOQTRGLGCIITSLGRDKNQVGEVQVSTATQ 1070
 QY 241 QSFPLATCVNVCWTFVHAGSKTLAPKPGITQMTNNVDOLVGMQAPGASMPCTGCG 300
 DB 1071 QSFPLATCVNVCWTFVHAGSKTLAPKPGITQMTNNVDOLVGMQAPGASMPCTGCG 1130
 QY 301 SLDYLVTRADVTPVRRGDSRGLSPRPVSYLKGSSGGLPLCSGNAVGFRAAVCTR 360

Db 1131 AVDLVITVMTADVPIVRKDDRGALLSPRLSTLKGSSGCVLGRHAGVLPRAAVCA 1190
 QY 361 RGVAAVDFIPVESMETTR 380
 Db 1191 RGVAKSIDIPVESLDAVTR 1210

RESULT 12

T08841
 polyprotein - douroucouli hepatitis GB virus A
 C/Species: douroucouli hepatitis GB virus A
 C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
 C/Accession: T08841
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A/Reference number: Z16486; MUID:98120818; PMID:9460920
 A/Accession: T08841
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 13005 <ERK>
 A/Cross-references: EMBL:AF023425; NID:g2828599; PIDN:AAC40502.1; PID:g2828600
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: polyprotein

Query Match 20.6%; Score 408.5; DB 2; Length 3005;
 Best Local Similarity 31.3%; Pred. No. 2.9e-25;
 Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

QY 41 RGRD-----AIIILTCVAPHELIPIITKLLAIFGLMLVQAGITVYVVAQGLIRAC 96
 Db 813 RGRDVRVTVAVVAVAGIIFREVVRCSALTA-----LAALLDSIDYLETL-ITLTA 864
 QY 97 MLVRKA-----AGHYVQMAFMKLAALTGYVVDHLPLQDMAHAGRLDAVAVEPY 148
 Db 865 QPAAALALDSLFLGLADITRAIFVRLEKRGVTLFPHCGQVXGAAAILXDLGVLEEV 924
 QY 149 IFSMEVKITWGDITACGDIISGLFVSARREIILG--PADNREGQGRLLAPITAY 206
 Db 925 SVIARDCVITRADARTLACQGRVGLPVARRGDEVAVGFPSRALPFGVPTAAPVV- 983
 QY 207 SQGTGILGCIITSLTRDXKQVGEVQVSTATQSLACVNGVCMVTHGSGSTIAG 266
 Db 984 MORLGFPSVYKTSMLRDRREHGSIVLGTSTRSMGTCVNGVMTTFHGSNAKTLAG 1043
 QY 267 PKGPIQMTYNVDODLVGMQAPFARSMPTCTGSSDLYVTRHADVIVRRRSGDSRGS 326
 Db 1044 PVGVNCRMSPSDDAVVPLPSASCLPECKCTGQVWCIRN--DGALCHGSLSKVEL 1101
 QY 327 LSPRPVSYLKGSGGGLLCPSGHAAGIFRAAVCTRGV-----AKVNDIPVES 374
 Db 1102 DLPTFISDFRGSSGSPILCDEGHVGMV-VSVLHRGVKVTGVRYVKEWETLPKDS 1155

RESULT 13

T08839
 polyprotein - marmoset hepatitis GB virus A
 C/Species: marmoset hepatitis GB virus A
 C/Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
 C/Accession: T08839
 R/Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A/Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A/Reference number: Z16486; MUID:98120818; PMID:9460920
 A/Accession: T08839
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: genomic RNA
 A/Residues: 1-2970 <ERK>
 A/Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AAC40501.1; PID:g2828598
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: polyprotein

Query Match 17.2%; Score 342.5; DB 2; Length 2970;

Best Local Similarity 28.9%; Pred. No. 9.1e-20;
 Matches 103; Conservative 56; Mismatches 133; Indels 65; Gaps 12;

QY 65 LLALIFG-----LMVLQAGITKVPYFRAQGLIRACMLVRKAGHYVQMAFMKLA 116
 Db 816 LVAAFMFVEIRIAVCAVAFILGFGFDDVDYILEVYVSSPVLRLAVLDSLVAAAGKLA 875
 QY 117 ALTGTYVDHLTPIOD--WAAG-----LRDLAVAVEPIFSMEVKITWGDPTA 165
 Db 876 T---TWLVEKLRKKNCFLYAAGQVTRTAQQLQMGFALPEVAVHEDCAMVDAARTL 932
 QY 166 ACGDIISGLPVSARGREIILGPDNREGQWR-----LAPITAVSQGTGILGCIIT 219
 Db 933 SCQGSVHGKPVVARGGEVLGLVNGV---WELPGEVTPAPVAVH-HHGKGFFGVVKT 987
 QY 220 SLTGRDXKQVGEVQVSTATQSLFATCVNGVCTVFGAGSKTLIAGPKPIQVYTNVD 279
 Db 988 SMTGMDIEHGNVYVLTSTRSMGTCVNGVMTTFHGSNAKTLAQMGPVNSRMSAS 1047
 QY 280 QDLVGMQAPFARSMPTCTGSSDLYVTRHADVIVRRRSGDSRGLS----- 328
 Db 1048 DDVAVYPLPVGAKCLEPCKCOPQGVWVI-----RND--GALCHGTLAGTVEJDL 1094
 QY 329 PRPVSYLKGSGGGLLCPSGHAAGIFRAAVCTRG-----VAKVNDIPVESMETT 378
 Db 1095 PAELCDFRGSSGSPILCDEGHVGMV-LSVLHRGSRVTGIRYTKPMETLPREAITHT 1150

RESULT 14

H71426
 hypothetical protein - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 A/Variety: Columbia
 C/Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
 R/Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirkx
 P.; Medier, H.; Medler, S.; Kotter, P.; Ertian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 Nature 391, 485-488, 1998
 A/Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,
 erioft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reichman, S.; Ansc
 C.; Chalmers, N.
 A/Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 A/Reference number: A71400; MUID:98121113; PMID:9461215
 A/Accession: H71426
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-692 <BEV>
 A/Cross-references: GB:Z97340; NID:g2244950; PID:e327492; PID:g2244965
 C/Genetics:
 A/Map position: 4C0P9-4G3845

Query Match 5.6%; Score 112; DB 2; Length 697;
 Best Local Similarity 24.1%; Pred. No. 0.25;
 Matches 97; Conservative 45; Mismatches 135; Indels 126; Gaps 23;

QY 13 LARLIWQLYITRFAHQLQVWIPPLVNRGGRDAIILLCAVHPELIPIITKLLAIFGP 72
 Db 83 IADLAGIYVLR-----QGNFPRASVYAGNCEIL-----KGPILMDLTFLR--FLT 131
 QY 73 LMYLQAGITKVPY--FVRAQGLIRACMLVR--KAAGHYVQMAFM----- 113
 Db 132 LCMV-----FSKPFPAVLEBAGYTHEDVLLQKPKAGGHTMOPAFITIRDTNSKICLLIR 188
 QY 114 -----KLAALTGTYVDHLTPIODMAHAGRLDAVAVEPIFSMEVKITWGDPTA 166
 Db 189 GHSIDQITPATGAIVPFRHSHVLDH--GGLSNVLTG-----AH 226
 QY 167 CGDIISGLPVSARGREIILGPDNREGQWR--RLAPITAVSQGTGILGCIITSLTGR 224
 Db 227 CG-----VVAAR-----WIAKLSVP-----CLTKXL----- 248
 QY 225 DKQVGEVQVNSTA-----TQSLFATCV-----NGVCMVTFHAGSKTLIAGPKGP--I 271

```

Db      249 DENP-SRKYQIVGHSLGGGTASLLTYILRQKFAATCFPAFGTPTNLMINGESGKFI 307
QY      272 TQMTYNDODLV--GWQAPPGASMTPTCGSSDLYLTRHADVIP-VRRGDSRGSLL 327
Db      308 TTIING--SDLVPTFSASSVDLRSEVTSSWSNDLRDQVHRTVLASVTKSATATGSR 365
QY      328 SPRPVSYLKSSSGGPLLCP--SGHAVGIFRAAVCTRGVAKAVD 368
Db      366 PSINAKAKVAGAGAILRPVSSGTQVAFLVNGC--GKIKCID 406

```

RESULT 15

```

VHMMH2
structural protein 2 precursor - hepatitis E virus (strain Burma)
C/Species: hepatitis E virus
C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Jul-1999
C/Accession: C40778
R/Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.
Virology 185, 120-131, 1991
A/Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi
A/Reference number: A40778; MUID:92024067; PMID:1926770
A/Accession: C40778
A/Molecule type: genomic RNA
A/Residues: 1-660 <TRM>
A/Cross-references: GB:W73218; NID:9330023; PIDN:AAA5736.1; PID:9330026
A/Note: the authors translated the codon Cgc for residue 2 as Ala
C/Superfamily: hepatitis E virus structural protein 2
C/Keywords: structural protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-660/Product: structural protein 2 #status predicted <SP2>

```

Query Match

```

Best Local Similarity 5.2%; Score 102.5; DB 1; Length 660;
Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

```

```

QY      63 TKLL--AIFGLMVLQAG-----ITKPYEYF--AQLIRACWLYKKAAGHYVQYA 111
Db      151 TNLVLAAPLSPLEPQDGNTHIMATASNVAQVVARATIRYRPLVPMNAVGYAISIS 210
QY      112 FKLALITGYVVDHLTPLODMAAGRLDAVAEPVIFSDMEVKIITWGADTAACGDI 171
Db      211 FWPOTTTPTSV-----DNNSITSTVRILVQPGIASSELYI----- 246
QY      172 SGLPVASRGRETIIGPAD--NFGQGWRLAPI--TAYSCQTRGL-----GCITSLTG 223
Db      247 -----PSERLHTRNQGMRASVETSGVAEEBATSGLVLCIHGSLVNSYTN 290
QY      224 -----RDKNQVEGEVQVNSTATQSF 244
Db      291 TPTGALGLDFALELEFRLNTPGNTNTRVSRYSTARRHLRAGDGTALTTTATRFM 350
QY      245 A---TCVAGV-----CMTVZH-----GAG-----SKTLAGPKG-PIT 272
Db      351 KDLFTSTNGVGEIGIALTLTFNLADTLGGLPTELLSSAGQLFYSRPVVSANGEPV 410
QY      273 QMTYNDODLVGWAQAPPGASMTPTCGSSDLYLV--TRHADVIPVRRGDSRG-SLLS 328
Db      411 KLYTSVEANA---QDQKGIAPHIDIDIGBSRVVIQDYDNQHEQDRPTSPASRPFVLR 466
QY      329 PRPVSYLK-----GSSGGLPCSGHAVGIFRAAVCTRGVAKAVDFIPV 372
Db      467 ANDVLMLSLTAAEYDOSTYGSSTGPVYV--SDSVTLVNVATGAQAVARSLDWTXY 519

```

Search completed: May 6, 2004, 09:37:18
 Job time : 13.3758 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 7.9319 Seconds
(without alignments)
2494.160 Million cell updates/sec

Title: US-10-650-585-12
Perfect score: 1987
Sequence: 1 ALLTSPYKVLARLIMW.....RGVAKAVDFIPVSMETMR 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1902	95.7	3010	1	POLG_HCVUA
2	1890	95.1	3010	1	POLG_HCVUT
3	1870	94.1	3010	1	POLG_HCVTW
4	1823	91.7	3010	1	POLG_HCVBK
5	1715	86.3	3011	1	POLG_HCVI
6	1703	85.7	3011	1	POLG_HCVH
7	1391	70.0	3033	1	POLG_HCVU8
8	1391	70.0	3033	1	POLG_HCVU6
9	102.5	5.2	660	1	VST2_HEVHU
10	102.5	5.2	660	1	VST2_HEVVA
11	101	5.1	564	1	SRS3C_ARATN
12	101	5.1	600	1	DP02_MOUSE
13	93.5	4.7	1380	1	CYAA_LEIPO
14	93	4.7	434	1	TOLB_CHLRE
15	92.5	4.7	706	1	TREB_HORSE
16	92	4.6	659	1	VST2_HEVME
17	91.5	4.6	485	1	VST2_HEVVA
18	91.5	4.6	660	1	VST2_HEVWY
19	91.5	4.6	3414	1	POLG_TBEVW
20	91	4.6	444	1	VA47_XYLFA
21	90.5	4.6	3414	1	POLG_TBEVH
22	89.5	4.5	961	1	ARCW_YERE
23	89	4.5	444	1	V327_XYFT
24	88.5	4.5	401	1	PILC_PSEBU
25	87.5	4.4	3412	1	POLG_TBSBU
26	86.5	4.4	347	1	MDHM_EUCU
27	86.5	4.4	470	1	MDHM_IAMM
28	86	4.3	338	1	GABE_NEICO
29	86	4.3	433	1	DCUA_WOLSU
30	85	4.3	441	1	MDHP_MESCR
31	85	4.3	470	1	NRAM_IATPA
32	85	4.3	730	1	H3LS_METNA
33	85	4.3	834	1	PWP2_SCHPO

RESULT 1
POLG_HCVUA STANDARD; PRT; 3010 AA.
ID POLG_HCVUA
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide (contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis B virus (HBV) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
DE Hepatitis C virus (isolate Japanese) (HCV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
OC NCBI_TaxID=11116;
PP [1]
RX MEDLINE=91089550; PubMed=2175903;
RA Kato N., Hijioka M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528 (1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849489;
RA Kato N., Hijioka M., Nakagawa M., Ootsuyama Y., Muraishi K., Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328 (1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate + (RNA) (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and that statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL, D90208, BA014233.1, -

DR PIR; A39253; GNNVCJ.
 DR HSSD; P26663; IUXP.
 DR MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR InterPro; IPR003903; Cys_Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_PS_PS.
 DR InterPro; IPR007094; RNA_pol_PSV1r.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR Pfam; PF0186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Polypeptin; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 95.7%; Score 1902; DB 1; Length 3010;

Best Local Similarity 94.2%; Pred. No. 6,9e-148;
 Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 2 ILTSPYKVLARLIMVQYLITREAHLOVMIPLNVGRDAILLTQAVHPELFD 61
 DB 828 ILTSPYKVFARLIMVQYLITREAHLOVMIPLNVGRDAILLTQAVHPELFD 687
 QY 62 ITKLLALFGPLMTQAGITVPRVRAQGLIRACMLVRAAGHYVOMAMKLAALTGT 121
 DB 888 ITKLLALFGPLMTQAGITVPRVRAQGLIRACMLVRAAGHYVOMAMKLAALTGT 947
 QY 122 YVYDHLTPLODMARAGLRLAVAVEPVIFSDMEVKIITWAGDTAACGDIISGLPVSAARG 181
 DB 948 YVYDHLTPLODMARAGLRLAVAVEPVIFSDMEVKIITWAGDTAACGDIISGLPVSAARG 1007
 QY 182 REILGPDNDEGGCWRLAITYASQCTRLGCIITSLTGRKNQVGEVQVSTATQ 241
 DB 1008 KEILGPDNDEGGCWRLAITYASQCTRLGCIITSLTGRKNQVGEVQVSTATQ 1067
 QY 242 SFATCVAGVCMVTFHAGSKTLAGPKPIQMTNTVDODLVGQAPRGASMPCTCGS 301
 DB 1068 SFATCVAGVCMVTFHAGSKTLAGPKPIQMTNTVDODLVGQAPRGASMPCTCGS 1127
 QY 302 SDLYLVTRHADVIFVRRRGRSGSLSPRPVSYLKSGSGGELLCPSCNAVGFRAVCTR 361
 DB 1128 SDLYLVTRHADVIFVRRRGRSGSLSPRPVSYLKSGSGGELLCPSCNAVGFRAVCTR 1187
 QY 362 GVAAVDPIPVESMETTR 380
 DB 1188 GVAAVDPIPVESMETTR 1206

RESULT 2
 ID POLG_HCVUT STANDARD; PRT; 3010 AA.
 AC 000259;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis virus)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_Taxid=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RT Virus Res. 23:39-53(1992).
 RL -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Glu in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins: a
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M84754; -; NOT_ANNOTATED_CDS.
 CC PIR, A40244; GNMVTW.
 CC PDB, 1N64; 25-FEB-03.
 CC PDB, 1N53; 08-APR-98.
 CC MEROPS, S29.001; -.
 CC MEROPS, U39.001; -.
 CC InterPro: IPR009003; Cys_Ser_tryp_sln.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_core.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_core.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4d.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_Rdrp.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR004109; Peptidase_C29.
 CC InterPro: IPR007095; RNA_pol_PS.
 CC InterPro: IPR007094; RNA_pol_PSV1r.
 CC Pfam: PF01543; HCV_core; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_core; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00271; Helicase_C; 1.
 CC Pfam: PF00938; Viral_Rdrp; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXDC; 1.
 CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 CC 3D-structure.
 CC INIT_MET 1 1
 CC -----
 CC REMOVED FROM CAPSID PROTEIN C BY THE
 CC CELLULAR AMINOPEPTIDASE.
 CC CORE PROTEIN (POTENTIAL).
 CC MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 CC NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 CC NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 CC PROTEASE/HELICASE NS3 (POTENTIAL).
 CC NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 CC NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 CC RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 CC POTENTIAL.
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC ATP (POTENTIAL).
 CC DECH BOX.
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 196 209
 CC CARBOHYD 209

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 94.1%; Score 1870; DB 1; Length 3010;
 Best Local Similarity 92.1%; Pred. No. 3e-145;
 Matches 349; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 2 LITLSPYKVLARLIWMLQYLITRVAHLQWIPPLNVRGGRDAIILTCVAPEDLFD 61
 DB LITLSPHKMFLALILWLVQYFIRBAHLQWIPPLNVRGGRDAIILTCVAPEDLFD 867
 QY 62 ITKLILAFGLPLVLOAGITKVPYFVPAQGLIRACMLVRAAGHYVOMAPKILALITGT 121
 DB ITKLILAFGLPLVLOAGITRIPYFVPAQGLIRACMLVRAAGHYVOMAPKILALITGT 947
 QY 122 VYVHPLFLOMAHAGRDIAVAEPIFSDMEKILITWGAADTAACGDIISGLPVSARRG 181
 DB VYVHPLFLOMAHAGRDIAVAEPIFSDMEKILITWGAADTAACGDIISGLPVSARRG 1007
 QY 182 REILGPADNEEGGKMLLAPITAYSOQTRGLGCIITSLTGRDKNVEGEVQVSTATQ 241
 DB REILGPADNEEGGKMLLAPITAYSOQTRGLGCIITSLTGRDKNVEGEVQVSTATQ 1067
 QY 242 SFLATCNGVCWTFPHAGSFTLAGPXPITOMWTANDOLVGMQAPRGASMTPTCCGS 301
 DB SFLATCNGVCWTFPHAGSFTLAGPXPITOMWTANDOLVGMQAPRGASMTPTCCGS 1127
 QY 302 SDLYLVRHADVIVRRRCDSRGLSPRPVSYLKGSGGPIILCPSSHAVGIFRAAVCTR 361
 DB SDLYLVRHADVIVRRRCDSRGLSPRPVSYLKGSGGPIILCPSSHAVGIFRAAVCTR 1187
 QY 1128 SDLYLVRHADVIVRRRCDSRGLSPRPVSYLKGSGGPIILCPSSHAVGIFRAAVCTR 1187
 DB 1188 GVAAKAVDFIPVESMETTR 380
 DB 1188 GVAAKAVDFIPVESMETTR 1206

RESULT 4
 POLG_HCVBK STANDARD; PRT; 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin).
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_Taxid=11105;
 RN [1]


```

FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT TURN 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT STRAND 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1176
FT TURN 1187 1188
FT STRAND 1189 1197
FT STRAND 1198 1202
FT TURN 1203 1204
FT STRAND 1204 1204
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFED9C CRC64;

Query Match 91.7%; Score 1823; DB 1; Length 3010;
Best Local Similarity 91.0%; Pred. No. 2,2e-141;
Matches 345; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 2 LITTSFYKYLRLRLVWLYLITRVANLQVWIPPLNVRGGDAIILLTCVHPELIRD 61
DB 828 LITTSFYKYLRLRLVWLYLITRVADLHWVIPPANRGRDAIILLTCVHPELIRD 887
QY 62 ITKLLAIFGLPLMWLQGITKVPYFVRAOGLIRACMLVRKAGGYVQNAFMKALVTGT 121
DB 888 ITKLLAIFGLPLMWLQGITKVPYFVRAOGLIRACMLVRKAGGYVQNAFMKALVTGT 947
QY 122 YVVDHLTPLODMAGLRDLAVAVEPITSDMEVKITITGADTACGDIISGLPVASARRG 181
DB 948 YVVDHLTPLODMAGLRDLAVAVEPITSDMEVKITITGADTACGDIISGLPVASARRG 1007
QY 182 REILLGADNPEEGQWELLAPITAYSCQTRGLGCIITSLTGDKKQVGEVQVSTARG 241
DB 1008 REILLGADNPEEGQWELLAPITAYSCQTRGLGCIITSLTGDKKQVGEVQVSTARG 1067
QY 242 SFLATCVNGVCMVTFVHAGSKTLAGEKPIITQVITVNDQDLVGMQAPFGARSMTPTCTGSS 301
DB 1068 SFLATCVNGVCMVTFVHAGSKTLAGEKPIITQVITVNDQDLVGMQAPFGARSMTPTCTGSS 1127
QY 302 SDLYLTVRRADVLPVRRRGDSGLSPRPVSTLKSSGCGPILLCPSGNAVGIFFRAVCTR 361
DB 1128 SDLYLTVRRADVLPVRRRGDSGLSPRPVSTLKSSGCGPILLCPSGNAVGIFFRAVCTR 1187
QY 362 GYAKAVDFIPVESMETTMR 380
DB 1188 GYAKAVDFIPVESMETTMR 1206

RESULT 5
POLG HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=91172826; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RT "Genetic organization and diversity of the hepatitis C virus.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DB EMBL; M62321; AAA5676.1; -.
DB PIR; A39166; GNMWC3.
DB PDB; 1AIV; 16-FEB-99.
DB PDB; 1HEI; 25-NOV-98.
DB MEROPS; S29.001; -.
DB MEROPS; U39.001; -.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; Peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PStir.
DR Pfam; PF01543; HCV_capsid_1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01533; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD180662; HCV_NS1; 1.
DR SMART; SM00487; DEADc; 1.
KM Polypeptide: Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KM Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KM 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006

```

REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CAPSID PROTEIN C (POTENTIAL).
 MATRIX PROTEIN (POTENTIAL).
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

```

FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1165 1167 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 1230 1237 ATP (POTENTIAL).
FT NP_BIND 1230 1237 DECH BOX.
FT SITE 1316 1319
FT CARBOHYD 1316 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FC5A99 CRC64;

Query Match 86.3%; Score 1715; DB 1; Length 3011;
Best Local Similarity 82.8%; Pred. No. 1,6e-132;
Matches 313; Conservative 32; Mismatches 33; Indels 0; Gaps 0;

```

```

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepadivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate H) (HCV).
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
DE Hepacivirus.
OC NCBI_TaxID=11108;
RN (1)
RN SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658600;
RX Inchauspe G., Zebedee S., Lee D.H.H., Sugtani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates."
RT Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).
RN (2)
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX MEDLINE=97331322; PubMed=9187654;
RX Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain."
RT Nat. Struct. Biol. 4:463-467 (1997).
RN (3)
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=98154321; PubMed=9493270;
RX Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding."
RT Structure 6:89-100 (1998).
RL (1) FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC (1) FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC (1) FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC (1) FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC (1) FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC (1) CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC (1) CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC (1) SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC (1) PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC (1) SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC (1) SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; M67463; AAA45534.1; -
DR PIR; A36814; GNWVCH.
DR PDB; 1HE1; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR INTERPRO; IPR009003; Cys_ser_trypsin.
DR INTERPRO; IPR001410; DEAD.
DR INTERPRO; IPR002522; HCV_capsid.

```


DB 1129 DLTAVTRHADVIFVRARRGDSGLSPRISLYKSGSGPLLCFCHAVGLFAAVCTFG 1188
 QY 363 VAKAVDFIPVSMETMR 380
 DB 1189 VAKAVDFIPVENLETMR 1206

RESULT 7
 POLG_HCVJ6 STANDARD; PRT: 3033 AA.

AC P26660;
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polypeptide [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP33); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
 DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate HC-J6) (HCV).
 OS Hepatitis C virus (isolate HC-J6) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11113;
 RX MEDLINE=2044440; Pubmed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kuzai K., Iizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.,
 RT Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions."
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL, D00944; BA00792.1; -
 CC F01, J01303; J01303.
 CC HSSP; P27958; 1HE1.
 CC MEROPS; S29.001; -
 CC InterPro: IPR0009003; Cys_Ser_trypsin.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002518; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR001490; HCV_NS4A.
 CC InterPro: IPR002868; HCV_NS5A.
 CC InterPro: IPR002166; HCV_RdRP.

DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4A; 1.
 DR Pfam: PF01001; HCV_NS4B; 1.
 DR Pfam: PF01506; HCV_NS5A; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT SITE 1234 1241
 FT SITE 1330 1333
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 224 224
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 FT SEQUENCE 3033 AA; 329165 MW; P957F5CIA273B95E CRC64;

Query Match 70.0%; Score 1391; DB 1; Length 3033;
 Best Local Similarity 66.0%; Pred. No. 7, 1e-106;
 Matches 250; Conservative 55; Mismatches 74; Indels 0; Gaps 0;

QY 2 LTLSPYVYVLLARLITWLTQYITRVEALQWIPPLNVRGGRDAIILLTCVHPELFD 61
 DB 832 LFTLTPGYKTLISRFIMWLCYLLITLAEAVQWAPPMQVGRDGIIMAVAFICGVVFD 891
 QY ITGLIAIFGPIMLVLOAGITKVPYFVRAQGLIRACGLYRKAGHYVQVAFKLAALTGT 121
 DB 892 ITKMLAVGPRAYLKGALTRPVPYFRAHALLRMCMTVRHLAAGRVQVLLALGRWTGT 951
 QY 122 YYYDHLTPLODVAHAGLRDLAAVEVPVFSDEVEKTIITGADTAACGDIISGLPVSARAG 181
 DB 952 YYYDHLTPMSDMAANGLRDLAAVAVEPIIPSPEKVIIVGAGTAACGDIILGLPVSARAG 1011

QY	182	REILILGPDNNEGOWMLLAPITAYASQOTRGLGIIISLGRDKNQVGGVQVASTQ	244
Db	1012	REVLILGPDNDGTSKMSLAPITAYAQOTRRLGLIVSMGRDKTEQAGELQVUSTYTQ	10712
QY	242	SFLATCVANGVCMYTFHGAAGXTLAGPKSPITOMYTNVDQILVGWQAPGASMTPTCTGS	301
Db	1072	SFLGTTISGVLMYTFHGAAGNXTLAGRSPVITOMYSSAEGDVLGWPSPPTGRSLRPTCGA	1131
QY	302	SDILVLRHAAVITFVRRRGDSRGLSLSRPVSYLKGSSGCPLLCPSGAAVGLFRAAVCTR	361
Db	1132	VDLVLVRNADVITFARRRGDRGALLSPRLTLKSGSGGEPVLCPRGHAIVGVFAAAVCSR	11912
QY	362	GVAAKADPEIPVESMETTMR	380
Db	1192	GVAKSIDFIPVETLDTYTR	1210

RESULT 8	
POLG_HCVJ8	
ID	POLG_HCVJ8
00000000	STANDARD;
00000000	PRT; 3033 AA

[illegible]

CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS2C and NS5 may play a role in the viral RNA replication.

CC - CATALYTIC ACTIVITY: N nucleoside triphosphate +
CC {RNA} (N).

CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipidprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way distorted, modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMEL; D10988; BAA01761.1; -
DR PIR; A40250; GNTVJ8.

DR	HSSP; P27958; LHEI.
DR	MEROPS; S29.001; -.

DR MEROPS; U39.001; -.
Tt+o+D+o; TDD0000003

DR ILIELPRO; LPR009003
DR InterPro; TBP001410

THE UNIVERSITY OF CHICAGO

DR	InterPro: IPR002522; HCV capsid.
DR	InterPro: IPR002521; HCV core.
DR	InterPro: IPR002519; HCV env.
DR	InterPro: IPR002531; HCV NS1.
DR	InterPro: IPR002518; HCV NS2.
DR	InterPro: IPR000745; HCV NS4a.
DR	InterPro: IPR001490; HCV NS4b.
DR	InterPro: IPR002868; HCV NS5a.
DR	InterPro: IPR002166; HCV RdRp.
DR	InterPro: IPR004109; Peptidase C29.
DR	InterPro: IPR007035; RNA pol D5 pb.
DR	InterPro: IPR007034; RNA pol_PSVit.
DR	Pfam: PFO1543; HCV_core; 1.
DR	Pfam: PFO1539; HCV_env; 1.
DR	Pfam: PFO1560; HCV_NS1; 1.
DR	Pfam: PFO1538; HCV_NS2; 1.
DR	Pfam: PFO2907; HCV_NS3; 1.
DR	Pfam: PFO1006; HCV_NS4a; 1.
DR	Pfam: PFO1001; HCV_NS4b; 1.
DR	Pfam: PFO1506; HCV_NS5a; 1.
DR	Pfam: PFO0998; Viral_RdRp; 1.
DR	ProDom: PD186062; HCV_NS1; 1.
DR	SMART: SM00487; DEXDc1; 1.
KW	Transmembrane; Coat protein; Envelope protein; Helicase; ATP-binding; protein; Hydrolyase; Setae; protease.
KW	Transmembrane; Nonstructural
FT	INIT_MET 1
FT	1
FT	CHAIN 1 115
FT	CHAIN 116 191
FT	CHAIN 192 383
FT	CHAIN 384 733
FT	CHAIN 734 1010
FT	CHAIN 1011 1619
FT	CHAIN 1620 2017
FT	CHAIN 1867 2016
FT	CHAIN 2018 3033
FT	CHAIN 347 359
FT	TRANSMEM 1087 1087
FT	ACT_SITE 1111 1111
FT	ACT_SITE 1169 1169
FT	NP_BIND 1234 1241
FT	1230 1323
FT	1323 1323
FT	CARBOHYD 196 196
FT	CARBOHYD 209 209
FT	CARBOHYD 233 233
FT	CARBOHYD 299 299
FT	CARBOHYD 305 305
FT	CARBOHYD 417 417
FT	CARBOHYD 423 423
FT	CARBOHYD 430 430
FT	CARBOHYD 448 448
FT	CARBOHYD 477 477
FT	CARBOHYD 534 534
FT	CARBOHYD 542 542
FT	CARBOHYD 558 558
FT	CARBOHYD 578 578
FT	CARBOHYD 627 627
FT	CARBOHYD 649 649
FT	CARBOHYD 1091 1091
FT	CARBOHYD 2038 2038
FT	CARBOHYD 2359 2359
FT	CARBOHYD 2811 2811
FT	3033 AA; 330177 NM; 1A175E3381FDDA CQC64;
QO	SEQUENCE

831 SFTPTTPAYKLLLSRYSVWMLSYMLVLAERQIQCNVPLPEYRGGRDGIWVAVILHPRIIVE 890

QY 61 DITKLLAIFGELMVLQAGITKVPYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTG 120
DB 891 EYTKMIALLGAVYLLKASLRIPIYVRHALLRVCTLVKHLAGARYICMLITIGRWTG 950
QY 121 TYVYDHLTPLOMAHAGLDLAVAVEPVFSDMEVKITWGDPAACGIIISGLVPSAR 180
DB 951 TYVYDHLTPLOMAHAGLDLAVAVEPVFSDMEVKITWGDPAACGIIISGLVPSAR 1010
QY 181 GRELLGPDNPFEGGMRLLAPITAYSCOOTRGLGCIITSLGARDNQVEGEVQVSTAT 240
DB 1011 GREVLGPDAGYTSKGMKLAIPITATYTOOTRGLGAIIVSLTRGRDNBQAGVQLSAYT 1070
QY 241 OSFLATCVNGVCTVFEHAGSKTLAPKSPITQMTYNNVDLIVGQAPPAASMTPTCTG 300
DB 1071 QFLGTSISGVLTWYHAGNKTLAGPKSPVQMTYSAEGDIVGWPSPGTXTSLDPCTCG 1130
QY 301 SSDLVYTRHADVTPRRRGDSRGSILSPRVSYLKSGSGGLCPGSHAVGIFRAAVCT 360
DB 1131 AVDLVYTRHADVTPRRRGDSRGSILSPRVSYLKSGSGGLCPGSHAVGIFRAAVCT 1190
QY 361 RGVAKAVDFIPVESMETMR 380
DB 1191 RGVAKSIDFIPVESLDAVTR 1210

RESULT 9

VS2 HEVBU STANDARD; PRT; 660 AA.

ID VS2 HEVBU STANDARD; PRT; 660 AA.
AC P29376;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
NCBI_TaxID=31767;
RN NCBI_TaxID=31767;
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RAY K.E., Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
full-length viral genome.";
RT Virology 185:120-131(1991).
RL -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M73218; AAA45736.1; -
DR PIR: C40778; VHWMH2.
DR InterPro: IPR004261; SP2.
DR InterPro: IPR008975; Viral_cap_coat.
DR Pfam: PF03014; SP2; 1.
DR Signal.
KM SIGNAL.
FT CHAIN 20 19 POTENTIAL.
FT SIGNAL 20 19 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCCA61C CRC64;

Query Match 5.2%; Score 102.5; DB 1; Length 660;
Best Local Similarity 19.8%; Pred. No. 0.93;
Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

QY 63 TKLL--AIFGLMVLQAG-----ITKVPYFVR--AQLIRACMLVRKAAGHYVQMA 111

DB 151 TNLVLYAPLSPLPLQDQNTTHIMTEASVAYQRYARATIRKRLVPAVAGYALIS 210
QY 112 FMKLAALTGIVYDHLTPLOMAHAGLDLAVAVEPVFSDMEVKITWGDPAACGII 171
DB 211 FWPQTTTTPSTV-----DMNSITSDVRLVOPGIASLTV----- 246
QY 172 SGLPVSARGREIILGPD--NFGGMRLLAPL-TVSCOOTRLL-----GCIITSLG 223
DB 247 -----PSERLHYNQMRSVETSGVAEEETSGVLMCTHGSIVNSYTN 290
QY 224 -----BDKNQVEGEVQVSTATOSFL 244
DB 291 TPTGALGLDPLALEPRNLTPTGNTNTRVSRVSTARRRRRADDGALETTTAATRFM 350
QY 245 A---TCVNGV-----CMTVFH-----GAG-----SKTLAGEKQ-PIT 272
DB 351 KDLYFTSTNGVGEIGRGIALTLFNLADTLGLPTLEISSAGGLFYSRPVSAANGEPTV 410
QY 273 QMTNVDLIVGQAPPAASMTPTCTGSSDLYV---TRHADVTPRRRGDSRG-SILS 328
DB 411 KLYTSVENA---QDDKIALPHDIDGESRVVQDYNQCHQDPRTPSPAPSPFVLR 466
QY 329 PRPVSYLK-----GSSGGLCPGSHAVGIFRAAVCTRGVAKAVDFIPV 372
DB 467 ANDVLMLSLTAEXDQSTYGSSTGPVY--SDSVTLNVVATGAQAVARSLDMTRY 519

RESULT 10

VS2 HEVPA STANDARD; PRT; 660 AA.

ID VS2 HEVPA STANDARD; PRT; 660 AA.
AC P33426;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Pakistan) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
NCBI_TaxID=33774;
RN NCBI_TaxID=33774;
RP SEQUENCE FROM N.A.
RX MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
RA Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
RL -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M80581; AAA5727.1; -
DR PIR: M80581; SP2.
DR InterPro: IPR004261; SP2.
DR InterPro: IPR008975; Viral_cap_coat.
DR Pfam: PF03014; SP2; 1.
DR Signal.
KM SIGNAL.
FT CHAIN 23 22 BY SIMILARITY.
FT SIGNAL 23 22 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70960 MW; 8085BC53CFB46FD3 CRC64;

Query Match 5.2%; Score 102.5; DB 1; Length 660;
Best Local Similarity 19.8%; Pred. No. 0.93;
Matches 82; Conservative 54; Mismatches 128; Indels 151; Gaps 20;

QY 63 TKLL--AIFGLMVLQAG-----ITKVPYFVR--AQLIRACMLVRKAAGHYVQMA 111

```

Dd 151 TNLVLAAPSPPLPDGCTNTHIMATASNAQYVARATITRPLVPAVAGVATISIS 210
Qy 112 FMKLAALTYGVYDHLTPLODMAGRLDVAVEPVISDMKVIITWGAADTAACGDI 171
Dd 211 FMYQTITPTSV-----DMNSITSDVIRLVGCIASELVI-----246
Qy 172 SGPLVARRRREILIGPAD--NFEQGGWRL-----APINAYSO 208
Dd 247 -----PSEKLRHQGWRKSVETSGVAEEBATSGLVLMCIHSFVNSYN 290
Qy 209 QT-RGLIGCT-----ITSIGRDKNQ-----VEGEVQVSTATOSPL 244
Dd 291 TPTYGALGLDFALELEFRNLTPGNTNTRVSRYSSTARHRLRGADGTALTTATATRM 350
Qy 245 A-----TCVNGV-----CWTVH-----GAG-----STLAGPKG-PIT 272
Dd 351 KDLYFTSTNGVSGIRGIALTLFNLADTLGLPTELISAGQLFYSSPVSANSEPV 410
Qy 273 QMTYNDQDVLVQWQAFGARSMPTCTGSSDLYLV---TRHADVIPRRRSGRG-SLTS 328
Dd 411 KLTYSVENA-----QOKGIALPHDIDLGESRVVIGQYDQHEQDRTPEPAPSRPSVLR 466
Qy 329 PRVYSYK-----GSSGGLPCPSGHAAGIFRAAVCTRGVAKAVDFIPV 372
Dd 467 ANDVLMSTLPAEYDOSTYGSSTGPPYV--SDSVTLVNVATGAQAVARSLDWTXV 519

RESULT 11
SRSC_ARATH
ID SRSC_ARATH STANDARD; PRT; 564 AA.
AC P37107; O82570;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Signal recognition particle 54 kDa protein, chloroplast.precursor
DE (SRP54) (54 chloroplast protein) (54CP) (PFC).
GN PFC OR AT5G03940 OR P8F6_150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, Columbia;
RX MEDLINE=94012817; PubMed=8408079;
RA Franklin A.E., Hoffman N.E.;
RT "Characterization of a chloroplast homologue of the 54-kDa subunit of
RT the signal recognition particle.";
RL J. Biol. Chem. 268:22173-22180(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;
RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc,
RT encoding the 54 kDa subunit of chloroplast signal recognition
RT particle.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, Columbia;
RX MEDLINE=21016721; PubMed=1130714;
RA Tabara S., Kaneko T., Nakamura Y., Koriaki H., Kato T., Asanizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara N., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naito K., Okumura S., Shinjo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoecking T., Pegin K., Spiehl J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Danc M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozerov P., Riley A., Strommatt C.,

```

```

RA Wagner-McPherson C., Kollan A., Yeakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez K., Hoon See L., Vli D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Matra M.A.,
RA Marijansen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Mambet R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Lanthan K.-D., Terryn N., Hartley N., Bert E., Johnson S.,
RA Lanthan S.-A., McCullach B., Robben J., Grymompier B., Zimmermann W.,
RA Rampeger U., Medler H., Balke K., Medler E., Peters S.,
RA van Steevenen T., Ditse W., Noeliman P., Klein Lankhorst R.,
RA Weltzeneger M., Bothe G., Rose M., Haut J., Bernieret S., Hempel S.,
RA Feldpasch M., Lamberth S., Villarroel R., Gielens J., Adiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K., Rüd S., Schoof H.,
RA Scheller C., Zaccaria P., Mewes H.-W., Beyan M., Franz P.,
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana".
RT Nature 408:823-826(2000).
CC -1- FUNCTION: May target chloroplast proteins to either the thylakoid
CC or envelope membranes.
CC -1- SUBCELLULAR LOCATION: Chloroplast stroma.
CC -1- TISSUE SPECIFICITY: Most abundant in green shoot tissue and
CC lower levels seen in the roots and etiolated buds.
CC -1- SIMILARITY: Belongs to the GTP-binding SRP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21970; CAA7981.1; -
CC EMBL; AF092168; AAC64139.1; -
CC EMBL; AL162873; CAB85514.1; -
CC PIR; S36637; S36637.
CC HSSP; O07347; 1PFH.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR000897; SRP54.
CC InterPro; IPR004125; SRP54_SBP.
CC InterPro; IPR004780; SRP_sub.
CC Pfam; PF00448; SRP54_1.
CC Pfam; PF02881; SRP54_N_1.
CC Pfam; PF02978; SRP_SBP_1.
CC SMART; SM00382; AAA_1.
CC TIGRFAMs; TIGR00959; ffh; 1.
CC PROSITE; PS00300; SRP54_1.
CC Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;
CC TRANSIT peptide.
CC KW CHAIN 1..75
CC FT 76 75 CHOROPLAST.
CC FT 76 564 SIGNAL RECOGNITION PARTICLE 54 KDA
CC FT DOMAIN 76 370 PROTEIN.
CC FT DOMAIN 371 564 G-DOMAIN.
CC FT NP_BIND 183 190 GTP (BY SIMILARITY).
CC FT NP_BIND 265 269 GTP (BY SIMILARITY).
CC FT NP_BIND 323 326 GTP (BY SIMILARITY).
CC FT CONFLICT 76 76 E -> V (IN REF. 2).
CC SQ SEQUENCE 564 AA; 61232 MW; 423F7285FB903E4 CRC64;

Query Match 5.1%; Score 101; DB 1; Length 564;
Best Local Similarity 26.1%; Pred. No. 1;
Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;

Qy 54 VHEPLI---FDITKLAALFGLPLVYLGAI-----TKYPTVRAQGLIRACMLVR 100
Dd 154 VHEPLVLMGVESELOFAKSGPTVILLGLQGVGKTVCAKLAICYLKKG--KSCMLI- 210
Qy 101 KAAAGHVQAFAFKLAAL---TGTYYDHLTPLO--DMAHAGRLDVAVEPVISDMKVI 155
Dd 211 --ADVTPRAIDVLYLGQGVVPTVYAGTDVAKDIAKQGLKAKK-----NNVDV 261
Qy 156 KIITWGAADTAACGDIISGVPVARRGRRL-----LGPANFEGGGMRLAPITAYSOOT 210

```

```

Db      262 VIN-----DTRGRLOIKKMMDELDKDYKFLNPFVILLVVDATGQ--EAAALVTFNVEI 315
Qy      211 RGLAGCITSLNCRDKNQVEGEVQVS 237
Db      316 -GITGAILTLTDGDSRGGAALSVKEVS 341

RESULT 12
DP02 MOUSE STANDARD; PRT; 600 AA.
ID      DP02 MOUSE
AC      P33611;
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
GN      POLA2.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.
RX      MEDLINE=93216788; PubMed=8463324;
RA      Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
RA      Hanaka F.;
RT      "Molecular cloning of the cDNAs for the four subunits of mouse DNA
RT      polymerase alpha-primase complex and their gene expression during
RT      cell proliferation and the cell cycle.";
RL      J. Biol. Chem. 268:8111-8122(1993).
CC      -1- FUNCTION: May play an essential role at the early stage of
CC      chromosomal DNA replication by coupling the polymerase
CC      alpha/primase complex to the cellular replication machinery (By
CC      similarity).
CC      -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme
CC      (subunits A, B, C and D), which is assembled throughout the cell
CC      cycle. The largest subunit (subunit A) has DNA polymerase
CC      activity, the two smallest subunits (subunits C and D) have DNA
CC      primase activity. Subunit B binds to subunit A.
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- PM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC      PHASE (BY SIMILARITY).
CC      -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL: D13546; BAA02746.1; -.
DR      PIR: B46642.
DR      MGI: 996690; PolA2.
DR      InterPro: IPR007200; DNA_pol_alpha_B.
DR      Pfam: PF04058; DNA_pol_alpha_B; 1.
KW      DNA replication; Nuclear protein; Phosphorylation.
FT      DOMAIN 101 107 POLY-GU.
FT      DOMAIN 115 157 PRO/SER/THR-RICH (HYDROPHILIC).
SQ      SEQUENCE 600 AA; 66267 MW; 79F9ABE6EF3FBC CRC64;

Query Match 5.1%; Score 101; DB 1; Length 600;
Best Local Similarity 24.8%; Pred. No. 1.1;
Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

```

```

Qy      185 ILGPADNREGGKRLIATAYSQOTRGLJCITSLTGRDKNQVEGEVQVSTATQSL 244
Db      145 LLSPS-----FSPSAPPSQK-----YTSRTN-----GEVTTFGSAQ--- 178
Qy      245 ATCVAGVCMTVFHAGAGSKTL--AGPKGPIQTQMYTNVDDPLVG 284
Db      179 -----GLSMGRGSGSVSLKVVGPDPRLTGSYKAMFQGLMG 215

RESULT 13
ID      CYAA LEIDO STANDARD; PRT; 1380 AA.
AC      Q27675;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Receptor-type adenylylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-
DE      lyase) (Adenylyl cyclase).
GN      PAC-A.
OS      Leishmania donovani.
OC      Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX      NCBI_TaxID=5661;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=IS Sudanese;
RX      MEDLINE=95340554; PubMed=7615561;
RA      Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.;
RT      "A family of putative receptor-adenylylate cyclases from Leishmania
RT      donovani.";
RL      J. Biol. Chem. 270:17551-17558(1995).
CC      -1- FUNCTION: Could act as a receptor for a unknown ligand.
CC      -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC      -1- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC      -1- SUBCELLULAR LOCATION: Membrane-bound.
CC      -1- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote)
CC      but not in the mammalian host stage of the parasite life cycle.
CC      -1- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL: U17042; AAA74998.1; -.
DR      PIR: T18309; T18309.
DR      InterPro: IPR001054; G_cyclase.
DR      Pfam: PF00211; guanylate_cyc; 1.
DR      SMART: SM00044; CYCC; 1.
DR      PROSITE: PS50125; GUANYLATE CYCLASES_2; 1.
KW      Lyase; CAMP biosynthesis; Transmembrane; Receptor; Glycoprotein;
KW      Metal-binding; Magnesium.
FT      DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 35 55 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 56 891 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 892 912 POTENTIAL.
FT      DOMAIN 913 1380 CYTOPLASMIC (POTENTIAL).
FT      METAL 938 938 MAGNESIUM (BY SIMILARITY).
FT      METAL 981 981 MAGNESIUM (BY SIMILARITY).
FT      CARBOHYD 422 432 N-LINKED (GLCNAC...) (POTENTIAL).
FT      CARBOHYD 478 478 N-LINKED (GLCNAC...) (POTENTIAL).
FT      CARBOHYD 497 497 N-LINKED (GLCNAC...) (POTENTIAL).
FT      CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL).
SQ      SEQUENCE 1380 AA; 151692 MW; 6B2D5F7D3C1107A0 CRC64;

Query Match 4.7%; Score 93.5; DB 1; Length 1380;
Best Local Similarity 21.0%; Pred. No. 13;
Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;

```

```

Db 101 GGRPIKIL-----HPPDDQDNLYDIIEVILHSLAROEKILAVGPIIDRLTAALSNADV 155
Qy 76 LQGITVP-----YFRAQGLIRACILYKAKAGHYV-----QMAFK 114
Db 156 VQSGMLIAFTSSGVRTSDSYFTRAEPMELKVLVLM-----HIVRLRRARVAEVR 210
Qy 115 LAAATG-----TYVYDHLTPLODWAHAGLRDLA-VEPVIFSDMEVK-----156
Db 211 ---LTGMHFGSEELTYQDTLTSL-----LRDPALVLYVPSSESVDEHAFDPAVAD 260
Qy 157 -----ITWGNADTAACDIIISGLPVGARRELLIG-----PADNFEQ 195
Db 261 TNPQVITVMAAPVQVYIFLEKVLTPRISAVVISCMTORVVPDYVELLSAGSIKQ 320
Qy 196 GWELLAPITAYSOOTRGLLCITTSITGRDKNOVEG-----VQVSTATOSFLATCVNGV 251
Db 321 DGRILASAT-----SPVSGELKYMELVLAQMSNIE-----NSG 356
Qy 252 CMTVFHAGSKTLAGEP-----GPITQMTYVVDODLVGMQAPPARSMPTCTCGSSDLYLV 307
Db 357 SFYVYDDSTETELGRARSEAPLSRKYT-VDE--FQAHPETAKLMALGMLSTGLVQ 412
Qy 308 TRADVIVPARRGDSKSLSPRPV---SYLKSSGSPILIC-PSGHAVGIFRAVCTRGV 363
Db 413 TLEQTMIVARSTYKAGLFGNQRFVIGGDIYLDYGGP--CEPLAGPLG--ASCYCNQGG 468
Qy 364 AKAV-----DFIPVESME 376
Db 469 HSSILTVLQNASWDIVDSSFK 490

RESULT 14
TOLE_CHLITE STANDARD; PRT; 434 AA.
AC Q8KED0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE TOLE protein precursor.
GN TOLE OR CT0636.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobiaceae; Chlorobi; Chlorobiales; Chlorobiaceae;
OC Chlorobiaceae; Chlorobi; Chlorobiales; Chlorobiaceae;
OC NCBI_Taxid=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JUS / ATCC 49652 / DSM 12025;
RX MEDLINE=2103685; Pubmed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolony J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Niernan W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Redune D.,
RA Vamathevan J., Knorr H., Bryant D.A., Fraser C.M.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.,
RT "The complete genome sequence of Chlorobium tepidum TJS, a
RT photosynthetic, anaerobic, green-sulfur bacterium."
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
CC -!- FUNCTION: Involved in the tonB-independent uptake of proteins (by
CC similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to the tole family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC EMBL; AE012837; AAM71875.1; -

```

```

DR TIGR; CT0636; -.
DR HAMAP; MP_00671; -.
KW Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 27 Potential.
FT CHAIN 28 434 TOL protein.
SQ SEQUENCE 434 AA; 47275 MW; FIA347B89C7A0F99 CRC64;

Query Match
Best Local Similarity 21.2%; Pred. No. 3.4;
Matches 69; Conservative 47; Mismatches 139; Indels 70; Gaps 14;

Qy 70 FGLPMVLAGITKYPYFRAQGLIRACMLYKAKAGHYVGMAFKALNGTYVYDHLTP 129
Db 8 FALCQIMFGMLVFPFLRAEE-VGEYIAIRK-EGASRIAVLDKISADGKQ-----58
Qy 130 LQDMAH-----GLRDLA-VEPVIFSDMEVKIITWGNADTAACDIIISGLPV 177
Db 59 -REWARSLDVTYTKNGLDFTGLFNLPA--PLNIRNGQNGGILNFASIVSGDIIYAGGSVT 115
Qy 178 ARRGRELI-----LGPADNFEQGRRLAPITAYSOQTRGLLCITTS 220
Db 116 KRSGRPVLEMHVYDSGKSLARTYTESQQLAIGLRFQADLVELLTKRSVFGTRIV 175
Qy 221 LTGRDKQ-----VEGEVQVSTATOSFLATCV-----NGVCVTFHAGSKTLAGEPKG 269
Db 176 VANTRKKEIMCDPDEENVQLTNSISILTPAVSPDGYIAMTV--TSGKNLYIKN 233
Qy 270 PITQMTYVVDODLV-----GMQAPPARSMPTCTC--GSSDLYVTRADVIVPARRGDSR 323
Db 234 IATKAKVSVNHGVCISPAWR--PGITVTLVTLSEGDODLYLI--RADGTVERRLTKGG 289
Qy 324 GSLSLP-----RPVSYLKSSGSP 342
Db 290 GIDVSPFSPGSGMAVSTRQGGP 314

RESULT 15
TRFE_HORSE STANDARD; PRT; 706 AA.
AC P27425;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sero transferrin precursor (Transferrin) (Betal-1-metal
DE binding globulin).
GN TF.
OS Equus caballus (Horse).
OC Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_Taxid=9796;
RX MEDLINE=93277958; Pubmed=8504171;
RA Carpenter M.A., Broad T.B.; transferrin."
RA "The cDNA sequence of horse transferrin."
RL Biochim. Biophys. Acta 1173:230-232 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Extraembryonic tissue;
RC McQuell K.J., Adams M.H., Baker C.B.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transferrins are iron binding transport proteins which
CC can bind two atoms of ferric iron in association with the binding
CC of an anion, usually bicarbonate. It is responsible for the
CC transport of iron from sites of absorption and heme degradation to
CC those of storage and utilization. Serum transferrin may also have
CC a further role in stimulating cell proliferation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -!- DOMAIN: Composed of two homologous domains.
CC -!- SIMILARITY: Belongs to the transferrin family.
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

DR EMBL; M69020; AAA30958.1; -;
 DR EMBL; U21127; AAA63684.1; -;
 DR PIR; S33761; S33761.
 DR HSP; P02787; 1A8R.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PRO0422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 706 SEROTRANSFERRIN.
 FT REPEAT 20 357 1.
 FT REPEAT 358 706 2.
 FT DISULFID 26 64 BY SIMILARITY.
 FT DISULFID 36 55 BY SIMILARITY.
 FT DISULFID 134 215 BY SIMILARITY.
 FT DISULFID 174 190 BY SIMILARITY.
 FT DISULFID 177 198 BY SIMILARITY.
 FT DISULFID 187 200 BY SIMILARITY.
 FT DISULFID 248 262 BY SIMILARITY.
 FT DISULFID 360 623 BY SIMILARITY.
 FT DISULFID 366 398 BY SIMILARITY.
 FT DISULFID 376 389 BY SIMILARITY.
 FT DISULFID 423 701 BY SIMILARITY.
 FT DISULFID 441 664 BY SIMILARITY.
 FT DISULFID 474 550 BY SIMILARITY.
 FT DISULFID 496 522 BY SIMILARITY.
 FT DISULFID 508 592 BY SIMILARITY.
 FT DISULFID 519 533 BY SIMILARITY.
 FT DISULFID 590 604 BY SIMILARITY.
 FT DISULFID 642 647 BY SIMILARITY.
 FT METAL 79 79 IRON 1 (BY SIMILARITY).
 FT METAL 111 111 IRON 1 (BY SIMILARITY).
 FT METAL 209 209 IRON 1 (BY SIMILARITY).
 FT METAL 270 270 IRON 1 (BY SIMILARITY).
 FT METAL 413 413 IRON 2 (BY SIMILARITY).
 FT METAL 449 449 IRON 2 (BY SIMILARITY).
 FT METAL 544 544 IRON 2 (BY SIMILARITY).
 FT METAL 612 612 IRON 2 (BY SIMILARITY).
 FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
 FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
 FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
 FT SIMILARITY).
 FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
 FT SIMILARITY).
 FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 FT SIMILARITY).
 FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 FT SIMILARITY).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SEQUENCE 706 AA; 78094 MM; 1A0FA566C0409D8A CRC64;

Query Match 4.7%; Score 92.5; DB 1; Length 706;

Best Local Similarity 21.5%; Pred. No. 6.7; Mismatches 109; Indels 81; Gaps 17;

DB 110 MAFKALALGTGVY---YDHITLPQDVAHGLRDLAVAVPEVIFSDMEVKITTWGA--- 162
 321 LGFRLRPAMDWTWYLVGYEYV-----AINRLREDIRPEVKD-ECKVKVCAIGHH 371

QY 163 DPAAGD--ITSGLPVSARRG-----ETLGPADNFEQGMRL-----LAPITAY 206
 DB 372 EKVCKDESVNVSNGNIECESAQSTEDCIAKIVKGADAMSIDGFIYIAGKGVPLAE 431
 QY 207 SQGTGGLGCIITSLTGDKQVEGEVGVSTANQSFLATCVNGVCWTFVHAGSKTLAG 266
 DB 432 NYETSGSACVTPREGYH-----AVAVKSSSDPDLT-----W-----NSLKG 470
 QY 267 PKGPIITQWTVNDQDLVGMQAPPGARSMTPTCGSSDYLVTNRADVIPRRRGDSRSL 326
 DB 471 KK-----SCHTGVDR-TAGNIPMG-----LYSEIKGCEPDKFFREGCAPGYR 513
 QY 327 LSPRVSYLKSGSGGP-LLC-PSGHA-----VGIFRAVCTRGVAKAVDFIPESME 376
 DB 514 RNSTLCNLCIGSASGPGRCEPNNHERYYGTGARFCLVEKGDA---FVHQTFE 566

Search completed: May 6, 2004, 09:31:52
 Job time : 9.93319 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 35.5407 seconds
(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-12
Perfect score: 1987
Sequence: 1 ALLTSPYKVLARLWL.....RGVAKAVDPVPSMETTR 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	1921	96.7	3010	12 Q9J3F9	Q9J3F9 hepatitis C
2	1918	96.5	3010	12 Q9J3F6	Q9J3F6 hepatitis C
3	1915	96.4	3010	12 Q9J3H7	Q9J3H7 hepatitis C
4	1909	96.1	3010	12 Q9J3D6	Q9J3D6 hepatitis C
5	1908	96.0	3010	12 Q9J3E4	Q9J3E4 hepatitis C
6	1908	96.0	3010	12 Q807P3	Q807P3 hepatitis C
7	1907	96.0	3010	12 Q9J3G6	Q9J3G6 hepatitis C
8	1905	95.9	3010	12 Q68788	Q68788 hepatitis C
9	1905	95.9	3010	12 Q9J3H3	Q9J3H3 hepatitis C
10	1905	95.9	3010	12 Q9J3D7	Q9J3D7 hepatitis C
11	1903	95.8	1186	12 Q81755	Q81755 hepatitis C
12	1903	95.8	1284	12 Q81817	Q81817 hepatitis C
13	1903	95.8	3010	12 P89966	P89966 hepatitis C
14	1902	95.7	3010	12 Q9J3H9	Q9J3H9 hepatitis C
15	1902	95.7	3010	12 Q9J3A2	Q9J3A2 hepatitis C
16	1902	95.7	3010	12 Q9J310	Q9J310 hepatitis C

17	1902	95.7	3010	12 Q9Q1X6	Q9Q1X6 hepatitis C
18	1902	95.7	3010	12 Q9Q1X5	Q9Q1X5 hepatitis C
19	1901	95.7	3008	12 Q9J3F4	Q9J3F4 hepatitis C
20	1900	95.6	3010	12 Q9Q1Y3	Q9Q1Y3 hepatitis C
21	1900	95.6	3010	12 P90191	P90191 hepatitis C
22	1899	95.6	3010	12 Q9J3H0	Q9J3H0 hepatitis C
23	1898	95.5	3010	12 Q81760	Q81760 hepatitis C
24	1898	95.5	3010	12 Q9Q1X8	Q9Q1X8 hepatitis C
25	1898	95.5	3010	12 Q9Q1X7	Q9Q1X7 hepatitis C
26	1896	95.4	3010	12 Q9J311	Q9J311 hepatitis C
27	1896	95.4	3010	12 Q9Q1Y5	Q9Q1Y5 hepatitis C
28	1895	95.4	3010	12 P88603	P88603 hepatitis C
29	1895	95.4	3010	12 Q9J3H5	Q9J3H5 hepatitis C
30	1895	95.4	3014	12 Q9J3H0	Q9J3H0 hepatitis C
31	1893	95.3	3010	12 Q9J3E9	Q9J3E9 hepatitis C
32	1892	95.2	3010	12 Q68826	Q68826 hepatitis C
33	1892	95.2	3010	12 Q9J3H6	Q9J3H6 hepatitis C
34	1892	95.2	3010	12 Q9J3F0	Q9J3F0 hepatitis C
35	1891	95.2	3010	12 Q9Q1Y4	Q9Q1Y4 hepatitis C
36	1889	95.1	3011	12 Q9J3E3	Q9J3E3 hepatitis C
37	1889	95.1	3015	12 Q9J3H5	Q9J3H5 hepatitis C
38	1887	95.0	3010	12 Q9J3H5	Q9J3H5 hepatitis C
39	1885	94.9	3010	12 Q02829	Q02829 hepatitis C
40	1885	94.9	3010	12 Q9Q1Y6	Q9Q1Y6 hepatitis C
41	1885	94.9	3010	12 Q09796	Q09796 hepatitis C
42	1884	94.8	3010	12 Q9J3E5	Q9J3E5 hepatitis C
43	1882	94.7	3010	12 Q9J3G3	Q9J3G3 hepatitis C
44	1881	94.7	1275	12 Q06642	Q06642 hepatitis C
45	1881	94.7	3010	12 Q9W1K8	Q9W1K8 hepatitis C

ALIGNMENTS

RESULT 1

Q9J3F9 PRELIMINARY; PRT: 3010 AA.

AC Q9J3F9; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitis C virus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=MD3;
RA Nagayama K., Kuroaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease progression."
RL Submitted (Nov-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).

DR EMBL AF207774; AA65964.1; --
DR F1R; A61196; A61196.
DR F1R; P00246; P00246.
DR F1R; P80329; P80329.
DR HSSP; P27958; 1HE1.
DR MEROPS; S29.001; --
DR MEROPS; U39.001; --
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008336; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006518; P:electron transport; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR InterPro; IPR002518; HCV_NS1.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVlr.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01506; HCV_NS4b; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 KM Coat protein: Envelope protein: Glycoprotein: Nonstructural protein:
 KM Polyprotein: RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327102 MW; 7162CDB93B5E0C7 CRC64;

Query Match 96.7%; Score 1921; DB 12; Length 3010;
 Best Local Similarity 95.0%; Pred. No. 5.5e-153;
 Matches 361; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIMVQYITREVALQVIMPLNVRGGRDIIILLTCAVPELIF 60
 DB 827 ALLTSPYKVFARLIMVQYITREVALHWPPLNVRGGRDIIILLTCAVPELIF 886
 QY 61 DIKLLAIIFGLPMVQAGITKVPYFVRAQGLIRACMLVRKXAGGYVOMAEKLAALTG 120
 DB 887 DIKLLAIIFGLPMVQAGITKVPYFVRAQGLIRACMLVRKXAGHYVOMAEKLGALTG 946
 QY 121 TVYVDHLTPLOMAHAGLDLAVAVPVTFSMEVKIITWGDYTAAGDIISGLPVSAR 180
 DB 947 TVYVDHLTPLOMAHAGLDLAVAVPVTFSMETKIIITWGDYTAAGDIISGLPVSAR 1006
 QY 181 GREIILGPADNFGGQWRLLAPITAYSQOTRGLIGIITSLGRDKNQVEGVVVSAT 240
 DB 1007 GREIILGPADNFGGQWRLLAPITAYSQOTRGLIGIITSLGRDKNQVEGVVVSAT 1066
 QY 241 QSTLATCVNGCVTFVHAGSKTLGPKGPITQMTYNTVDVLVQAPPGASMTPTCTG 300
 DB 1067 QSTLATCVNGCVTFVHAGSKTLGPKGPITQMTYNTVDVLVQAPPGASMTPTCTG 1126
 QY 301 SSGIYVTHHADVTPRRGDSGSLSPRPVSYLKGSSGGLLCPSGHVGIFPAAVCT 360
 DB 1127 SSGIYVTHHADVTPRRGDSGSLSPRPVSYLKGSSGGLLCPSGHVGIFPAAVCT 1186
 QY 361 RGVAKAVDFIPVESMETWR 380
 DB 1187 RGVAKAVDFIPVESMETWR 1206

RESULT 2
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.
 ID Q9DTE6
 AC Q9DTE6
 DT 01-VAR-2001 (TrEMBLrel. 16, Created)
 DT 01-VAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepaciviruses.
 OK NCBI_TaxID=11103;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HCV1142;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Hachihara T., Ohta Y., Kanai K., Mario H., Baba K., Hijioka M.,
 RA Masuko S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 with hepatocellular carcinoma: the 'progression score' revisited.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC CC
 CC EMBL; AB049091; BAB1804.1; .
 DR PIR; A61196; A61196.
 DR PIR; PS0329; PS0329.
 DR HSSP; P2663; LUXP.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003958; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR GO; GO:0009003; Cys Ser trypsin.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVlr.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01506; HCV_NS4b; 1.
 DR Pfam; PF00271; Helicase_C; 1.

DR Pfam; PF00998; Viral RdRp; 1.
 DR Prodom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC1; 1.
 DR SMART; SM00490; HELICC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydroxylase; Nonstructural protein; Polyprotein;
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 96.5%; Score 1918; DB 12; Length 3010;
 Best Local Similarity 95.0%; Pred. No. 9.9e-153;
 Matches 360; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMWLQYLITRVAHLQWIPPLNVRGRDAIILLTCAVHDLIFD 61
 DB LTLSPYKVLARLIMWLQYLITRVAHLQWIPPLNVRGRDAIILLTCAVHDLIFD 887
 QY 62 ITKLIAIGPLMTLOAGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
 DB ITKLIAIGPLMTLOAGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 947
 QY 122 YVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 181
 DB YVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 1007
 QY 948 YVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 1007
 DB YVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 1007
 QY 182 REILGAPADNFBEGQWRLLAPITAYSQQTRGLIGCIITSLTGRDKXQVEGVQVSTATQ 241
 DB REILGAPADNFBEGQWRLLAPITAYSQQTRGLIGCIITSLTGRDKXQVEGVQVSTATQ 1067
 QY 242 SFPLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 301
 DB SFPLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 1127
 QY 1068 SFPLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 1127
 DB SFPLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 1127
 QY 302 SDLYLTVRHADVIPVRRGDSRGLSLSPREVSYLKSSGGPILCPSGHVAVGIFRAAVCTR 361
 DB SDLYLTVRHADVIPVRRGDSRGLSLSPREVSYLKSSGGPILCPSGHVAVGIFRAAVCTR 1187
 QY 1128 SDLYLTVRHADVIPVRRGDSRGLSLSPREVSYLKSSGGPILCPSGHVAVGIFRAAVCTR 1187
 DB SDLYLTVRHADVIPVRRGDSRGLSLSPREVSYLKSSGGPILCPSGHVAVGIFRAAVCTR 1187
 QY 362 GVAKAVDFIPVBSMETTR 360
 DB GVAKAVDFIPVBSMETTR 1206

RESULT 3
 Q9J3H7 PRELIMINARY; PRT; 3010 AA.

ID Q9J3H7
 AC Q9J3H7;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD15;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 RT progression.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL; AF207756; AAF65946.1; -;
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P00804; P00804.
 DR PIR; P00329; P00329.
 DR HSP; P26663; IJXP.
 DR GO; GO:0016021; C:Integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0003968; F:RNA binding; IEA.
 DR GO; GO:0003973; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-cysteine peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR GO; GO:0009003; Cys_ser_cypsin.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR Prodom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC1; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327365 MW; D8653F7317FFA106 CRC64;

Query Match 96.4%; Score 1915; DB 12; Length 3010;
 Best Local Similarity 94.5%; Pred. No. 1.8e-152;
 Matches 359; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIMWLQYLITRVAHLQWIPPLNVRGRDAIILLTCAVHDLIF 60
 DB ALLTSPYKVLARLIMWLQYLITRVAHLQWIPPLNVRGRDAIILLTCAVHDLIF 886
 QY 61 DITKLIAIGPLMTLOAGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 120
 DB DITKLIAIGPLMTLOAGITKVPYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 946
 QY 121 TVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 180
 DB TVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 1006
 QY 947 TVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 1006
 DB TVYDHLTPLODMWAGLRDLAAVEEVIKIDMEVKIITWGADTAAAGDIISGLPVSARG 1006
 QY 181 GREIILGAPADNFBEGQWRLLAPITAYSQQTRGLIGCIITSLTGRDKXQVEGVQVSTAT 240
 DB GREIILGAPADNFBEGQWRLLAPITAYSQQTRGLIGCIITSLTGRDKXQVEGVQVSTAT 1066
 QY 1007 GREIILGAPADNFBEGQWRLLAPITAYSQQTRGLIGCIITSLTGRDKXQVEGVQVSTAT 1066
 DB GREIILGAPADNFBEGQWRLLAPITAYSQQTRGLIGCIITSLTGRDKXQVEGVQVSTAT 1066
 QY 241 OSFLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 300
 DB OSFLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 1126
 QY 1067 OSFLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 1126
 DB OSFLATCVNGCVMTVEFHAGSKTLGPKGPIITQMTNVDQDLVGMQAPPGASMTPTCTGS 1126
 QY 301 SSDLYLTVRHADVIPVRRGDSRGLSLSPREVSYLKSSGGPILCPSGHVAVGIFRAAVCTR 360
 DB SSDLYLTVRHADVIPVRRGDSRGLSLSPREVSYLKSSGGPILCPSGHVAVGIFRAAVCTR 360

Db 1127 SSDLVTRHADVIYVRRRDSRGSLLSPRVSLKSSGGPLCPGSHAVGIFRAAVCT 1186
 QY 361 RGVAKAVDFIVESMETTR 380
 Db 1187 RGVAKAVDFIVESMETTR 1206

RESULT 4
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.
 AC Q9DTE6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV221;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Hatahara T., Ohta Y., Kanai K., Marto H., Baba K., Hijioka M.,
 RA Mishiuro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 RT with hepatocellular carcinoma: the 'progression score' revisited.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL: AB049101; BAB1814.1; -;
 DR PIR: A61196; A61196.
 DR PIR: P00246; P00246.
 DR PIR: PS0329; PS0329.
 DR HSP: P26663; LUXP.
 DR GO: GO:0016821; C: integral to membrane; IEA.
 DR GO: GO:0019028; C: viral capsid; IEA.
 DR GO: GO:0019031; C: viral envelope; IEA.
 DR GO: GO:0005524; F: ATP binding; IEA.
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F: electron transporter activity; IEA.
 DR GO: GO:0016787; F: hydrolase activity; IEA.
 DR GO: GO:0003723; F: RNA binding; IEA.
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F: structural molecule activity; IEA.
 DR GO: GO:0016740; F: transferase activity; IEA.
 DR GO: GO:0006118; F: electron transport; IEA.
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; F: transcription; IEA.
 DR GO: GO:0019079; F: viral genome replication; IEA.
 DR GO: GO:0019087; F: viral genome transformation; IEA.
 DR InterPro: IPR009003; Cys Ser. trypsin.
 DR InterPro: IPR000345; Cys Ser. trypsin.
 DR InterPro: IPR001410; DEAD_heme_BS.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA pol. D5. PS.
 DR InterPro: IPR007094; RNA pol. PSVIR.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.

DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF0186062; HCV_NS1; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolase; Nonstructural protein; Polypeptide;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327108 MW; D8182D810F78BE4 CRC64;

Query Match 96.1%; Score 1909; DB 12; Length 3010;
 Best local similarity 95.3%; Pred. No. 5,7e-152;
 Matches 361; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 2 LLLTSPYKVLARLWMLQVITRVEAHQVWITPINTVGRSDAIIITCAVHPETIFD 61
 Db 828 LLLTSPYKVLARLWMLQVITRVEAHQVWITPINTVGRSDAIIITCAVHPETIFD 867
 QY 62 ITKLILAFGRLVNLQAGITKVPYFVRAQGLIRACMLVRKAGGHVYQVAFMKLAALVGT 121
 Db 888 ITKLILAFGRLVNLQAGITKVPYFVRAQGLIRACMLVRKAGGHVYQVAFMKLAALVGT 947
 QY 122 VYVHPLPLDPMKAGRLDAVAVEPIESDMVKIITWADPAAAGDITSGIPVARRG 181
 Db 948 VYVHPLPLDPMKAGRLDAVAVEPIESDMVKIITWADPAAAGDITSGIPVARRG 1007
 QY 182 REILLGPADEFGGKMLLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVOSTATQ 241
 Db 1008 REILLGPADEFGGKMLLAPITAYSOQTRGLGCIITSLTGRDKNOVEGEVOSTATQ 1067
 QY 242 SFLATCVGCVWTFYFHAGSKTLAGEPKGPTTQMTYTNVDQDLVWQAPPGARSTPTCTCGS 301
 Db 1068 SFLATCVGCVWTFYFHAGSKTLAGEPKGPTTQMTYTNVDQDLVWQAPPGARSTPTCTCGS 1127
 QY 302 SDDLTVTRHADVIYVRRRDSRGSLLSPRVSLKSSGGPLCPGSHAVGIFRAAVCTR 361
 Db 1128 SDDLTVTRHADVIYVRRRDSRGSLLSPRVSLKSSGGPLCPGSHAVGIFRAAVCTR 1187
 QY 362 GVAKAVDFIVESMETTR 380
 Db 1188 GVAKAVDFIVESMETTR 1206

RESULT 5
 Q9DTE4 PRELIMINARY; PRT; 3010 AA.
 AC Q9DTE4;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV150;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Hatahara T., Ohta Y., Kanai K., Marto H., Baba K., Hijioka M.,
 RA Mishiuro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 RT with hepatocellular carcinoma: the 'progression score' revisited.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL, AB048093, BAB18806.1, --
 DR PIR, A61196, A61196.
 DR PIR, PQ0246, PQ0246.
 DR PIR, PQ0804, PQ0804.
 DR PIR, PS0329, PS0329.
 DR HSPD, P26633, IJXP.
 DR GO, GO:0016021, C:integral to membrane, IEA.
 DR GO, GO:0018028, C:viral capsid, IEA.
 DR GO, GO:0019031, C:viral envelope, IEA.
 DR GO, GO:0005524, F:ATP binding, IEA.
 DR GO, GO:0008026, F:ATP dependent helicase activity, IEA.
 DR GO, GO:0005489, F:electron transporter activity, IEA.
 DR GO, GO:0003723, F:RNA binding, IEA.
 DR GO, GO:0003968, F:RNA-directed RNA polymerase activity, IEA.
 DR GO, GO:0008238, F:serine-type peptidase activity, IEA.
 DR GO, GO:0005198, F:structural molecule activity, IEA.
 DR GO, GO:0016740, F:transferase activity, IEA.
 DR GO, GO:0006118, P:electron transport, IEA.
 DR GO, GO:0006508, P:proteolysis and peptidolysis, IEA.
 DR GO, GO:0006350, P:transcription, IEA.
 DR GO, GO:0018079, P:viral genome replication, IEA.
 DR GO, GO:0015087, P:viral transformation, IEA.
 DR InterPro, IPR009003, Cys_Ser_typsin.
 DR InterPro, IPR001410, DEAD.
 DR InterPro, IPR002522, HCV capsid.
 DR InterPro, IPR002521, HCV core.
 DR InterPro, IPR002519, HCV env.
 DR InterPro, IPR002531, HCV NS1.
 DR InterPro, IPR002518, HCV NS2.
 DR InterPro, IPR000745, HCV NS4a.
 DR InterPro, IPR001490, HCV NS4b.
 DR InterPro, IPR002868, HCV NS5a.
 DR InterPro, IPR002166, HCV RdRp.
 DR InterPro, IPR001650, Helicase_C.
 DR InterPro, IPR004109, Peptidase_C29.
 DR InterPro, IPR007095, RNA_pol_DS_PS.
 DR InterPro, IPR007094, RNA_pol_PSVir.
 DR Pfam, PF01543, HCV capsid, 1.
 DR Pfam, PF01542, HCV core, 1.
 DR Pfam, PF01539, HCV env, 1.
 DR Pfam, PF01560, HCV NS1, 1.
 DR Pfam, PF01538, HCV NS2, 1.
 DR Pfam, PF02907, HCV NS3, 1.
 DR Pfam, PF01006, HCV NS4a, 1.
 DR Pfam, PF01506, HCV NS4b, 1.
 DR Pfam, PF01506, HCV NS5a, 1.
 DR Pfam, PF00271, Helicase_C, 1.
 DR Pfam, PF00998, Viral_RdRp, 1.
 DR ProDom, PD186062, HCV_NS1, 1.
 DR SMART, SM00487, DEXDC1, 1.
 DR PROSITE, PS00190, CYTOCHROME C, 1.
 DR Coat protein, Envelope protein, Glycoprotein, Nonstructural protein, RNA polymerase, RNA-directed RNA polymerase, Transmembrane, Transmembrane.
 SO SEQUENCE 3010 AA, 327324 MW, 3DB6CF249BD1151C CMC64.

Query Match 96.08, Score 1908, DB 12, Length 3010;
 Best Local Similarity 94.58, Pred. No. 6.9e-152;
 Matches 359, Conservative 10, Mismatches 11, Indels 0, Gaps 0;

QY 1 ALLTSLSYVLLARILWVLYITRVEAHLOWVPIPLNRRGGSDAIIITTCVHPELLF 60
 DB 827 ALLTSLSYVLLARILWVLYITRVEAHLOWVPIPLNRRGGSDAIIITTCVHPELLF 86
 QY 61 DITKLLAIFGLPLVLTAGITKVPYFVRAOGLIRACMLVKAAGHYVQAEMKLAALTG 120
 DB 887 DITKLLAIFGLPLVLTAGITKVPYFVRAOGLIRACMLVKAAGHYVQAEMKLAALTG 946
 QY 121 TTYVVDHLLPLQDMAHAGRLDAVVEPVISDMEVKIITGADPAACGDIISGLPVSAAR 180
 DB 947 TTYVVDHLLPLQDMAHAGRLDAVVEPVISDMEVKIITGADPAACGDIISGLPVSAAR 1006

QY 181 GREILLGPADNFEQCGKRLIAPITANSQOTRGLIGCTIISLRDKNQVGEVQVSTAT 240
 DB 1007 GREILLGPADNFEQCGKRLIAPITANSQOTRGLIGCTIISLRDKNQVGEVQVSTAT 1066
 QY 241 QSFATCGNVCMTVPHGASXTLAGPKGPIITQMTNVDDLVGMQAPFGASMTPTCTG 300
 DB 1067 QSFATCGNVCMTVPHGASXTLAGPKGPIITQMTNVDDLVGMQAPFGASMTPTCTG 1126
 QY 301 SSDLYLTRHADYIPVRRDSDSGSLISPPVSYLKGSSGCPILCPGGAHVGFRAVCT 360
 DB 1127 SSDLYLTRHADYIPVRRDSDSGSLISPPVSYLKGSSGCPILCPGGAHVGFRAVCT 1186
 QY 361 RGAKAVDFIPVESMETMR 360
 DB 1187 RGAKAVDFIPVESMETMR 1206

RESULT 6
 ID 0807P3 PRELIMINARY; PRT; 3010 AA.
 AC 0807P3;
 DT 01-JUN-2003 (TREMBlrel, 24, Created)
 DT 01-JUN-2003 (TREMBlrel, 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel, 25, Last annotation update)
 DE Polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MILE;
 RX MEDLINE=22047193, PubMed=12051758;
 RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
 Nio Y., Hosaka M., Miyamori Y., Shimotohno K.;
 RT "Subgenomic replicon derived from a cell line infected with the
 hepatitis C virus."
 RL Biochem. Biophys. Res. Commun. 293:993-999(2002).
 DR EMBL, AB080299, BACS4896.1, --
 DR GO, GO:0019028, C:viral capsid, IEA.
 DR GO, GO:0019031, C:viral envelope, IEA.
 DR GO, GO:0005524, F:ATP binding, IEA.
 DR GO, GO:0008026, F:ATP dependent helicase activity, IEA.
 DR GO, GO:0005489, F:electron transporter activity, IEA.
 DR GO, GO:0003723, F:RNA binding, IEA.
 DR GO, GO:0003968, F:RNA-directed RNA polymerase activity, IEA.
 DR GO, GO:0008238, F:serine-type peptidase activity, IEA.
 DR GO, GO:0005198, F:structural molecule activity, IEA.
 DR GO, GO:0016118, P:electron transport, IEA.
 DR GO, GO:0006508, P:proteolysis and peptidolysis, IEA.
 DR GO, GO:0006350, P:transcription, IEA.
 DR GO, GO:0019079, P:viral genome replication, IEA.
 DR GO, GO:0018087, P:viral transformation, IEA.
 DR InterPro, IPR009003, Cys_Ser_typsin.
 DR InterPro, IPR000345, Cys_heme_BS.
 DR InterPro, IPR001410, DEAD.
 DR InterPro, IPR002522, HCV capsid.
 DR InterPro, IPR002521, HCV core.
 DR InterPro, IPR002519, HCV env.
 DR InterPro, IPR002531, HCV NS1.
 DR InterPro, IPR002518, HCV NS2.
 DR InterPro, IPR000745, HCV NS4a.
 DR InterPro, IPR001490, HCV NS4b.
 DR InterPro, IPR002868, HCV NS5a.
 DR InterPro, IPR002166, HCV RdRp.
 DR InterPro, IPR001650, Helicase_C.
 DR InterPro, IPR004109, Peptidase_C29.
 DR InterPro, IPR007095, RNA_pol_DS_PS.
 DR InterPro, IPR007094, RNA_pol_PSVir.
 DR Pfam, PF01543, HCV capsid, 1.
 DR Pfam, PF01542, HCV core, 1.
 DR Pfam, PF01539, HCV env, 1.
 DR Pfam, PF01560, HCV NS1, 1.

DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELICG; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Polyprotein; 3010 AA; 327097 MW; E86418C7A723E686 CRC64;
 SQ SEQUENCE

Query Match 96.0%; Score 1908; DB 12; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 6,9e-152;
 Matches 360; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 ALLTSSYYVVLARLIWLYITRVEAHLYWIPPLNVRGRDAIILLTCAVPELIF 60
 DB 827 AFLTSPYYVFLARLIWLYITRVEAHLYWIPPLNVRGRDAIILLTCAVPELIF 886

QY 61 DITKLLAIFGPMVLAQAGITKPYFVRAGGIRACMLVRKKAAGHYVOMAFVKLAALNG 120
 DB 887 DITKLLAIFGPMVLAQAGITKPYFVRAGGIRACMLVRKKAAGHYVOMAFVKLAALNG 946

QY 121 TYYVDHLTPIQDMAHAGLDLAVAVPVI FSDMEVKIITWGADTAAAGDIISGLPVSAAR 180
 DB 947 TYYVDHLTPIQDMAHAGLDLAVAVPVI FSDMEVKIITWGADTAAAGDIISGLPVSAAR 1006

QY 181 GREILLGPADNFEQGCGWRLLAPITAYSOQTRGLGCIITSLGRDNQVEGSEVQVYSTAT 240
 DB 1007 GREILLGPADNFEQGCGWRLLAPITAYSOQTRGLGCIITSLGRDNQVEGSEVQVYSTAT 1066

QY 241 QSEFLATCVGVCWTFVHGAGSKTLAAGPKGPITQMTYVNDQDVLGMAAPGARSMTPTCG 300
 DB 1067 QSEFLATCVGVCWTFVHGAGSKTLAAGPKGPITQMTYVNDQDVLGMAAPGARSMTPTCG 1126

QY 301 SSRLYLVTRHADVTPRRRGDSGSLISRPVSYLKGSSGSGPLCPSGHAAVIFPAAVCT 360
 DB 1127 SSRLYLVTRHADVTPRRRGDSGSLISRPVSYLKGSSGSGPLCPSGHAAVIFPAAVCT 1186

QY 361 RGVAKAVDFIPVESMETTKR 380
 DB 1187 RGVAKAVDFIPVESMETTKR 1206

RESULT 7
 Q9J3G6 PRELIMINARY; PRT; 3010 AA.
 AC Q9J3G6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=MD26;
 RC Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 progression."
 RL Submitted (Nov-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND RNA POLYMERASE E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC EMBL; AF207767; AAF65957.1; -
 DR EMBL; AF207767; AAF65957.1; -
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.

DR PIR; P00254; P00254.
 DR PIR; PS0329; PS0329.
 DR HSS3; P26653; 1UXP.
 DR MEROPS; S29.002; -
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0019031; C:Viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006118; F:electron transport; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; F:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys Ser CysSta.
 DR InterPro; IPR001450; CysC_heme_BS.
 DR InterPro; IPR001410; DRAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA pol DS_PS.
 DR InterPro; IPR007094; RNA pol PSv1r.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Cost protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327165 MW; 74FAB6B80F24837B CRC64;

Query Match 96.0%; Score 1907; DB 12; Length 3010;
 Best Local Similarity 95.0%; Pred. No. 8,4e-152;
 Matches 361; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 ALLTSSYYVVLARLIWLYITRVEAHLYWIPPLNVRGRDAIILLTCAVPELIF 60
 DB 827 AFLTSPYYVFLARLIWLYITRVEAHLYWIPPLNVRGRDAIILLTCAVPELIF 886

QY 61 DITKLLAIFGPMVLAQAGITKPYFVRAGGIRACMLVRKKAAGHYVOMAFVKLAALNG 120
 DB 887 DITKLLAIFGPMVLAQAGITKPYFVRAGGIRACMLVRKKAAGHYVOMAFVKLAALNG 946

QY 121 TYYVDHLTPIQDMAHAGLDLAVAVPVI FSDMEVKIITWGADTAAAGDIISGLPVSAAR 180
 DB 947 TYYVDHLTPIQDMAHAGLDLAVAVPVI FSDMEVKIITWGADTAAAGDIISGLPVSAAR 1006

QY 181 GREILLGPADNFEQGCGWRLLAPITAYSOQTRGLGCIITSLGRDNQVEGSEVQVYSTAT 240
 DB 1007 GREILLGPADNFEQGCGWRLLAPITAYSOQTRGLGCIITSLGRDNQVEGSEVQVYSTAT 1066

QY 241 QSEIATCVAGVCTVTHGAGSKTLAGPKPITOMTNTVNDQDLVGNQAPPGASMPCTCG 300
 DB 1067 QSEIATCVAGVCTVTHGAGSKTLAGPKPITOMTNTVNDQDLVGNQAPPGASMPCTCG 1126
 QY 301 SSDLVVTRHADVIPIRRRGDSRGSILSPRPVSYLKSGSGGFLCPGSHAVGIFRAAVCT 360
 DB 1127 SSDLVVTRHADVIPIRRRGDSRGSILSPRPVSYLKSGSGGFLCPGSHAVGIFRAAVCT 1186
 QY 361 GYAKAVDPVPSMETTMR 380
 DB 1187 GYAKAVDPVPSMETTMR 1206

RESULT 8
 ID 068788 PRELIMINARY; PRT; 3010 AA.
 AC 068788;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE HCV polyprotein (Genome polyprotein).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96362158; PubMed=8720135;
 RA Seki M., Honda Y.;
 RT Phosphorolactate antisense oligodeoxynucleotides capable of
 RT inhibiting Hepatitis C virus gene expression: in vitro translation
 RT assay.";
 RT J. Biochem. 118:1199-1204 (1995).
 RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; DA5172; BAA08120.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; PS0329; PS0329.
 DR HSP; P26663; LUXP.
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: RNA dependent helicase activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:000368; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P: transcription; IEA.
 DR GO; GO:0019079; P: viral genome replication; IEA.
 DR GO; GO:0019087; P: viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_tyrpsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4A.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C9.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVtr.
 DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4A; 1.
 DR Pfam; PF01001; HCV_NS4B; 1.
 DR Pfam; PF01506; HCV_NS5A; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NSI; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326880 NM; EED840E6A050E766 CRC64;

Query Match 95.9%; Score 1905; DB 12; Length 3010;
 Best Local Similarity 94.5%; Pred. No. 1.2e-151;
 Matches 358; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 2 I L T S P Y K V L A R I L W L Q Y L I T R E A H L Q W I P L N V R G R D A I I L T C A V H P E L I F D 61
 DB 828 I L T S P Y K V L A K I L W L Q Y L I T R A E A H L Q W I P L N V R G R D A I I L T C A V H P E L I F D 887
 QY 62 I T K L L A F G P L M L O G I T V P F V R A O G I R A C M V R K A G G H Y V M A F K L A A L I G T 121
 DB 888 I T K L L A L L G P L M V Q A G I T R P F V R A O G I R A C M L V R K A G G H Y V M A F K L A A L I G T 947
 QY 122 Y V Y D H L T P L Q D M A H A G R D L A V A E P V I F S D M E V K I I T G A D T A C G D I I S G L P V S A R G 181
 DB 948 Y I Y D H L T P L R M A H A G R D L A V A E P V I F S D M E K I I T W G A D T A C G D I I I G L P V S A R G 1007
 QY 182 R E I L L G P A D N E G G C M L A R I T A Y S Q T R E L I G C I I T S L T G R K N Q V E G V Q V A S T A Q 241
 DB 1008 K E I L L G P A D S L E G G C M L A R I T A Y S Q T R E L I G C I I T S L T G R D N Q V E G V Q V A S T A Q 1067
 QY 242 S F I A T C V A G V C T V T H G A G S K T L A G P K P I T O M T N T V N D Q D L V G N Q A P P G A S M P C T C G S 301
 DB 1068 S F I A T C V A G V C T V T H G A G S K T L A G P K P I T O M T N T V N D Q D L V G N Q A P P G A S M P C T C G S 1127
 QY 302 S D L V V T R H A D V I P V R R R G D S R G S I L S P R P V S Y L K S G S G G F L C P S G H A V G I F R A A V C T R 361
 DB 1128 S D L V V T R H A D V I P V R R R G D S R G S I L S P R P V S Y L K S G S G G F L C P S G H A V G I F R A A V C T R 1187
 QY 362 G Y A K A V D P V P S M E T T M R 380
 DB 1188 G Y A K A V D P V P S M E T T M R 1206

RESULT 9
 ID 0933H3 PRELIMINARY; PRT; 3010 AA.
 AC 0933H3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD19;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 RT progression.";
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; AF207760; AA65950.1; -.

DR PIR; A61196; A61196.
 DR PIR; PS0329; PS0329.
 DR HSSP; P26663; IJXP.
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F: electron transporter activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P: transcription; IEA.
 DR GO; GO:0019079; P: viral genome replication; IEA.
 DR GO; GO:0019087; P: viral transformation; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid_1.
 DR Pfam; PF01542; HCV_core_1.
 DR Pfam; PF01539; HCV_env_1.
 DR Pfam; PF01560; HCV_NS1_1.
 DR Pfam; PF01538; HCV_NS2_1.
 DR Pfam; PF02907; HCV_NS3_1.
 DR Pfam; PF01006; HCV_NS4a_1.
 DR Pfam; PF01001; HCV_NS4b_1.
 DR Pfam; PF01506; HCV_NS5a_1.
 DR Pfam; PF00271; helicase_C_1.
 DR Pfam; PF00998; viral_RdRp_1.
 DR Pfam; PF016062; HCV_NS1_1.
 DR ProDom; PD186062; HCV_NS1_1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein.
 DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 DR SEQUENCE 3010 AA; 327234 MW; 44C34677649CB8DD CRC64;

Query Match 95.9%; Score 1905; DB 12; Length 3010;
 Best Local Similarity 93.7%; Pred. No. 1.2e-151;
 Matches 356; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIMWLQVLRVSHLQVMTPLNVRGGRALLLTCAVPELIF 60
 DB 827 ALLTSPYKVLARLIMWLQVLRVSHLQVMTPLNVRGGRALLLTCAVPELIF 886
 QY 61 DITKLLALFGLPMLVLAQGITVFPVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTG 120
 DB 887 SITKLLALFGLPMLVLAQGITVFPVFAOGLIRACMLVRKAAGHYVQMAFMKLAALTG 946
 QY 121 TVVYDHLTLQMAHAGLADLVAVEPVIFSDMEVKITITWGDPTAACGDIISGLPVSAAR 180
 DB 947 TVVYDHLTLQMAHAGLADLVAVEPVIFSDMEVKITITWGDPTAACGDIISGLPVSAAR 1006
 QY 181 GREILLGPADNPEGQGMRLIAPITAVSOOTRGLICITSLTGRDNQVEGSEVQVSTAT 240
 DB 1007 GREILLGPADNPEGQGMRLIAPITAVSOOTRGLICITSLTGRDNQVEGSEVQVSTAT 1066

QY 241 QSLATCVNGVCTVFPAGSKTLGAPKPTQMTYTNVDDLVGMQAPPGARSMTPTCTCG 300
 DB 1067 QSLATCVNGVCTVFPAGSKTLGAPKPTQMTYTNVDDLVGMQAPPGARSMTPTCTCG 1126
 QY 301 SSDLYLVTRADVPVRRRGGDSGLISPPVSTLKSSGGPLICPGHAGVIFRAAVCT 360
 DB 1127 SSDLYLVTRADVPVRRRGGDSGLISPPVSTLKSSGGPLICPGHAGVIFRAAVCT 1186
 QY 361 RGVAKAVDFVPESEMTTMR 380
 DB 1187 RGVAKAVDFVPESEMTTMR 1206

RESULT 10
 ID 09PDT7 PRELIMINARY; PRT; 3010 AA.
 AC 09PDT7.
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-OCT-2003 (Tremblrel. 16, Last sequence update)
 DE Genome polypeptin.
 OS Hepatitis C virus.
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11103;
 CC (1)
 CC SEQUENCE FROM N.A.
 RP STRAIN=HCV217;
 RA Hatahara T., Ohka Y., Kanai K., Maruo H., Baba K., Hijikata M.,
 RA Michio S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 RT with hepatocellular carcinoma: the 'progression score' revisited.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC
 DR EMBL; AB049100; BAB18613.1; -.
 DR PIR; A61196; A61196.
 DR PIR; PQ0804; PQ0804.
 DR PIR; PS0329; PS0329.
 DR HSSP; P26663; IJXP.
 DR MEROPS; S29_002; -.
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F: electron transporter activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006118; F: electron transport; IEA.
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P: transcription; IEA.
 DR GO; GO:0019079; P: viral genome replication; IEA.
 DR GO; GO:0019087; P: viral transformation; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR000345; CytC_heme_BS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF000271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR Pfam; PF186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polypeptide; RNA-directed RNA polymerase; Transferrase; Transmembrane;
 SO POLYPEPTIDE 3010 AA; 326909 MW; 550562EB0DA0519 CRC64;

Query Match 95.9%; Score 1905; DB 12; Length 3010;
 Best Local Similarity 94.5%; Pred. No. 1.2e-151;
 Matches 359; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

QY 1 ALTTSPYKVLARLIMVLYLTVREAHLOVWIPPLNVGGRAIILTCVHPELIF 60
 DB 827 ALTTSPYKVLARLIMVLYLTVREAHLOVWIPPLNVGGRAIILTCVHPELIF 886
 QY 61 DITKLLAIFGPMVLVQAGITKVPYFRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120
 DB 887 DITKLLAIFGPMVLVQAGITKVPYFRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 946
 QY 121 TTYVDHLTPLOMAHAGLRLAVAVPVFSDMEVYLLTMWADTACGGIISGLPSAR 180
 DB 947 TTYVDHLTPLOMAHAGLRLAVAVPVFSDMEVYLLTMWADTACGGIISGLPSAR 1006
 QY 181 GREILLGPADNFEQGMRLAPITAYSOQTRGLIGLITSLTRDRNQVEGEVQVSTAT 240
 DB 1007 GREILLGPADNFEQGMRLAPITAYSOQTRGLIGLITSLTRDRNQVEGEVQVSTAT 1066
 QY 241 QSLPLATCVGVCWTVTHGAGSKTLAPKGPITOMYNNVDLGVQAPPGASMTPTCTG 300
 DB 1067 QSLPLATCVGVCWTVTHGAGSKTLAPKGPITOMYNNVDLGVQAPPGASMTPTCTG 1126
 QY 301 SSPLVYVTRADYIPVRRRDSRGLSPRPVSYLLSGSSGGLPCSGHVGIFRAVCT 360
 DB 1127 SSPLVYVTRADYIPVRRRDSRGLSPRPVSYLLSGSSGGLPCSGHVGIFRAVCT 1186
 QY 361 RGVAKAVDFIPVESMETTKR 380
 DB 1187 RGVAKAVDFIPVESMETTKR 1206

RESULT 11
 Q81755 PRELIMINARY; PRT; 1186 AA.
 AC Q81755;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Polypeptide (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from Japanese
 patients with non-A, non-B hepatitis.";

RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S., Sugiyama Y., Kural K., Iizuka H., Machida A.,
 Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
 from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takemizawa A., Mori C., Manabe S., Murakami S., Fujita J., Onishi E.,
 Andoh T., Yoshida I., Okayama H.;
 RT "The structure and organization of the Hepatitis C virus genome
 RT isolated from human carriers.";
 RL J. Virol. 65:1105-1113(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C.,
 Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P., Lin M., Tai K., Liu P., Lin C., Chen D.;
 RT "The Taiwanese hepatitis C virus genome: Sequence determination and
 RT mapping the 5' termini of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kural K., Okada S., Yamamoto K., Iizuka H., Tanaka T.,
 Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: Comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 RN [7]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93323208; PubMed=8392606;
 RA Hijikata M., Mizushima H., Akagi T., Mori S., Kakiuchi N., Kato N.,
 Tanaka T., Kimura K., Shimotohno K.;
 RT "Two distinct proteinase activities required for the processing of a
 RT putative nonstructural precursor protein of hepatitis C virus.";
 RL J. Virol. 67:4665-4675(1993).
 RN [8]
 RP SEQUENCE FROM N.A.
 RA Hijikata M.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D11397; BAA20975.1; --
 DR PIR; A61196; A61196.
 DR PIR; PS0329; PS0329.
 DR PDB; 1DXP; 28-MAR-02.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0006506; F:proteolysis and peptidolysis; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys_Ser_cysteine.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR SMART; SM00487; DEXDC; 1.
 KM ATP-binding; Helicase; Hydrolyase.
 FT NON_TER
 SQ SEQUENCE 1186 AA; 126280 MW; 34170478BA23729A CRC64;
 Query Match 95.8%; Score 1903; DB 12; Length 1186;
 Best Local Similarity 94.7%; Pred. No. 5.4e-152;
 Matches 359; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 2 LTLSPYKVLRLIWMLOYITRVEAHLOVWIPPLNRRGRDAIILITCAVHPELLFD 61
 DB 106 LTLSPYKVLRLIWMLOYITRVEAHLOVWIPPLNRRGRDAIILITCAVHPELLFD 165
 QY 62 ITKLAIIFGLPLVNLQAGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
 DB 166 ITKLAIIFGLPLVNLQAGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 225
 QY 122 YVYDHLTPLODMAHAGLRDLAVALVEPVIFSDMEVKITWGADTAAAGDIISGLPVASARG 181
 DB 226 YVYDHLTPLODMAHAGLRDLAVALVEPVIFSDMEVKITWGADTAAAGDIISGLPVASARG 285
 QY 182 RETILGPANPESQGRRLIAPITAYSOQTRHLLGCIITSLNGRDNVGEVQVSTATQ 241
 DB 286 RETILGPANPESQGRRLIAPITAYSOQTRHLLGCIITSLNGRDNVGEVQVSTATQ 345
 QY 242 SPLATCVNGVCMVTFVAGSXTLAGKPIITQMTNVDDLVGWAQPPARSMPTCTCGS 301
 DB 346 SPLATCVNGVCMVTFVAGSXTLAGKPIITQMTNVDDLVGWAQPPARSMPTCTCGS 405
 QY 302 SDLYIVTRADYIPVARRDSRGSLSRPFSYLVKSSGSGPILCSGHAAGCFRAVCTR 361
 DB 406 SDLYIVTRADYIPVARRDSRGSLSRPFSYLVKSSGSGPILCSGHAAGCFRAVCTR 465
 QY 362 GVAKAVDFIPVESMETTMR 380
 DB 466 GVAKAVDFIPVESMETTMR 484

RESULT 12
 ID 081817 PRELIMINARY; PRT; 2284 AA.
 AC 081817;
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Polypeptide precursor (Genome polypeptide).
 OS Hepatitis C virus type 2.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=40271;
 RN 11)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94068484; PubMed=7504283;
 RA Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,
 Akagi T., Kimura K., Shimotohno K.;
 RT "Proteolytic processing and membrane association of putative
 nonstructural proteins of hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).
 RN 12)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94333810; PubMed=8056334;
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
 RT "Identification of the domain required for trans-cleavage activity of
 hepatitis C viral serine proteinase.";
 RL Gene 145:215-219(1994).
 RN 13)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95056078; PubMed=7966638;
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;

RT "Hepatitis C virus polypeptide processing: kinetics and mutagenic
 RT analysis of serine proteinase-dependent cleavage.";
 RL J. Virol. 68:8418-8422(1994).
 RN 14)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95156583; PubMed=7853491;
 RA Tanji Y., Hijikata M., Sato S., Kaneko T., Shimotohno K.;
 RT "Hepatitis C virus-encoded nonstructural protein NS4A has versatile
 RT functions in viral protein processing.";
 RL J. Virol. 69:1575-1581(1995).
 DR EMBL; D16435; BA03905.1; -
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P00246; P00246.
 DR PIR; P00246; P00246.
 DR HSP; P26663; IXP.
 DR GO; GO:0019012; C:Virus; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008226; F:serine-type peptidase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR00410; DEAD.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR004108; Peptidase C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Nonstructural protein; Polypeptide; RNA-directed RNA polymerase;
 KW Signal; Transferase.
 FT SIGNAL 4 20
 FT CHAIN 55 83
 FT CHAIN 21 54
 FT CHAIN 84 300
 FT CHAIN 301 931
 FT CHAIN 932 985
 FT CHAIN 986 1246
 FT CHAIN 1247 1693
 FT CHAIN 1694 2284
 SQ SEQUENCE 2284 AA; 247213 MW; DC272A1517046337 CRC64;
 Query Match 95.8%; Score 1903; DB 12; Length 2284;
 Best Local Similarity 94.7%; Pred. No. 1.3e-151;
 Matches 359; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 2 LTLSPYKVLRLIWMLOYITRVEAHLOVWIPPLNRRGRDAIILITCAVHPELLFD 61
 DB 106 LTLSPYKVLRLIWMLOYITRVEAHLOVWIPPLNRRGRDAIILITCAVHPELLFD 161
 QY 62 ITKLAIIFGLPLVNLQAGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
 DB 166 ITKLAIIFGLPLVNLQAGITKVFYFVRAOGLIRACMLVRKAAGHYVQMAFMKLAALTGT 221
 QY 122 YVYDHLTPLODMAHAGLRDLAVALVEPVIFSDMEVKITWGADTAAAGDIISGLPVASARG 181
 DB 226 YVYDHLTPLODMAHAGLRDLAVALVEPVIFSDMEVKITWGADTAAAGDIISGLPVASARG 281

QY 182 REILLGADNFEQGWMLLAPITVYSGQTRGLGCIITSLTGDKNQVGEVQVSTAQ 241
 Db 282 KEILLGADNFEQGWMLLAPITVYSGQTRGLGCIITSLTGDKNQVGEVQVSTAQ 341
 QY 242 SFLATCVGVCWTVYHAGSKITLAPRGKPTTQWYTNVDDLVGMQAPPGARSMTPCTCGS 301
 Db 342 SFLATCVGVCWTVYHAGSKITLAPRGKPTTQWYTNVDDLVGMQAPPGARSMTPCTCGS 401
 QY 302 SDLVLTVRHADVIVRRRGRSGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361
 Db 402 SDLVLTVRHADVIVRRRGRSGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 461
 QY 362 GVAKAVDFIPVESMETTMR 380
 Db 462 GVAKAVDFIPVESMETTMR 480

RESULT 13

P89366 PRELIMINARY; PRT; 3010 AA.

AC P89366; 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE RNA for polypyrroline (Genome polypyrroline).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBL_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=type 1b;
 RA Tanaka T.;
 RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=type 1b;
 RA TANAKA T.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 CC -1 SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; D89872; BAAL4035.1; -.
 DR PIR; A61196; A61196.
 DR PIR; F00246; F00246.
 DR PIR; F00804; F00804.
 DR PIR; P50329; P50329.
 DR HSP; P26663; IUXP.
 DR GO; GO:0019021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:0008236; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; F: transcription; IEA.
 DR GO; GO:0019079; F: viral genome replication; IEA.
 DR GO; GO:0019087; F: viral transformation; IEA.
 DR InterPro; IPR009003; Cys_Ser_Lysase.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR007451; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR004109; Peptidase C29.
 DR InterPro; IPR007095; RNA pol DS_PS.
 DR InterPro; IPR007094; RNA_pol_PstI_r.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polypyrroline; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327023 MW; E075BD9CFD8D1261 CAC64;

Query Match 95.8%; Score 1903; DB 12; Length 3010;
 Best local similarity 94.7%; Pred. No. 1.8e-151;
 Matches 359; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 2 LLTSPYKYLRLRLWLYLITRVBAHQVWIPPLNVRGGDDATILTCVHPPLIPD 61
 Db 828 LLTSPYKYLRLRLWLYLITRVBAHQVWIPPLNVRGGDDATILTCVHPPLIPD 887
 QY 62 ITKLLALFGLPLWYLGQITKVPYFVRAQGLIRACWVRAAGHYVQAFMKALITGT 121
 Db 888 ITKLLALFGLPLWYLGQITKVPYFVRAQGLIRACWVRAAGHYVQAFMKALITGT 947
 QY 122 YVVDHLPLODMWAGAGRDIAVEPVISDMVKIITGADTPAACDITSGIPVSRAG 181
 Db 948 YVVDHLPLODMWAGAGRDIAVEPVISDMVKIITGADTPAACDITSGIPVSRAG 1007
 QY 182 REILLGADNFEQGWMLLAPITVYSGQTRGLGCIITSLTGDKNQVGEVQVSTAQ 241
 Db 1008 KEILLGADNFEQGWMLLAPITVYSGQTRGLGCIITSLTGDKNQVGEVQVSTAQ 1067
 QY 242 SFLATCVGVCWTVYHAGSKITLAPRGKPTTQWYTNVDDLVGMQAPPGARSMTPCTCGS 301
 Db 1068 SFLATCVGVCWTVYHAGSKITLAPRGKPTTQWYTNVDDLVGMQAPPGARSMTPCTCGS 1127
 QY 302 SDLVLTVRHADVIVRRRGRSGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361
 Db 1128 SDLVLTVRHADVIVRRRGRSGSLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 1187
 QY 362 GVAKAVDFIPVESMETTMR 380
 Db 1188 GVAKAVDFIPVESMETTMR 1206

RESULT 14

Q933H9 PRELIMINARY; PRT; 3010 AA.

AC Q933H9; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Genome polypyrroline.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBL_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD13;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;

RT "Characteristics of hepatitis C viral genome associated with disease progression."

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL; AF207754; AAF65944.1; -

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; P00329; P00329.

DR HSSP; P26663; 1LXP.

DR GO; GO:0016021; C:Integral to membrane; IEA.

DR GO; GO:0019028; C:Viral capsid; IEA.

DR GO; GO:0019031; C:Viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F:electon transporter activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003686; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006118; F:electron transport; IEA.

DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; F:transcription; IEA.

DR GO; GO:0019079; F:vital genome replication; IEA.

DR GO; GO:0019087; F:vital transformation; IEA.

DR InterPro; IPR009003; Cys_Ser_tryp_sln.

DR InterPro; IPR000345; CytC_heme_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002531; HCV_env.

DR InterPro; IPR002531; HCV_NSI.

DR InterPro; IPR002518; HCV_NSI.

DR InterPro; IPR000745; HCV_NS4a.

DR InterPro; IPR001490; HCV_NS4b.

DR InterPro; IPR002868; HCV_NS5a.

DR InterPro; IPR002166; HCV_RdRp.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; Peptidase_C29.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NSI; 1.

DR Pfam; PF01538; HCV_NSI; 1.

DR Pfam; PF02907; HCV_NS3; 1.

DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.

DR Pfam; PF000721; Helicase_C; 1.

DR Pfam; PF009981; Viral_RdRp; 1.

DR Pfam; PF0186062; HCV_NSI; 1.

DR SMART; SMO0487; DEXDC; 1.

DR PRODom; PD186062; HCV_NSI; 1.

DR PROSITE; PS00190; CYTOCHROME C; 1.

KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.

KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.

SEQUENCE 3010 AA; 326984 MW; AFI2CC00E0A8B078 CR64;

Query Match 95.7%; Score 1902; DB 12; Length 3010;

Best local similarity 93.7%; Pred. No. 2.2e-151;

Matches 356; Conservative 11; Mismatches 13; Indels 0; Gaps 0;

QY 1 ALTTSPYKYLRLIMWLQYLITRVGAHQLQWIPPLNVRGGDAIILITCAVHPILIF 60

DB 827 ALTTSPYKYLRLIMWLQYLITRVGAHQLQWIPPLNVRGGDAIILITCAVHPILIF 866

QY 61 DITKLLAIFFPLNWLQAGITKVFYFVRAQGLIRACMLVRKAGHYVQMAFMKLAITG 120

DB 887 DITKLLAIFFPLNWLQAGITKVFYFVRAQGLIRACMLVRKAGHYVQMAFMKLAITG 946

QY 121 TVVYHHTLPLDMMAGLRDLAVAVEPIESDMEXKITTKGADTRAGCDITISGLPVARR 180

DB 947 TVVYHHTLPLDMMAGLRDLAVAVEPIESDMEXKITTKGADTRAGCDITISGLPVARR 1006

QY 181 GREILLGADNFEQGMRLAPITAYSQOTGLGCIITTSIGRDKQVEGEVQVSTAT 240

DB 1007 GREILLGADNFEQGMRLAPITAYSQOTGLGCIITTSIGRDKQVEGEVQVSTAT 1066

QY 241 QSLFATCNGVCMVFHFGAGSKITLAGPKPTQNTVNDQVLGMQAPPGARSXTPCTCG 300

DB 1067 QSLFATCNGVCMVFHFGAGSKITLAGPKPTQNTVNDQVLGMQAPPGARSXTPCTCG 1126

QY 301 SSDLYLTVRADVTPVRRGDSRLSPRVSYLKGSSGGLPCSGHAWGIFRAVCT 360

DB 1127 SSDLYLTVRADVTPVRRGDSRLSPRVSYLKGSSGGLPCSGHAWGIFRAVCT 1186

QY 361 RGVAKAVDFIPVESMETTMR 380

DB 1187 RGVAKAVDFIPVESMETTMR 1206

RESULT 15

Q99AU2 PRELIMINARY; PRT; 3010 AA.

AC Q99AU2;

DT 01-JUN-2001 (TREMblrel. 17, Created)

DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Genome polypeptin.

OS Hepatitis C virus type 1b.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

ON NCBI_Taxid=31647;

RX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=chimera of HCV-BK;

RA Thomson M., Nascentiment M., Gonzales S., Murthy K., Rehmann B., Liang J.

RT "Analyses of viral sequences and virus-specific immune responses during serial passage of an infectious hepatitis C virus serotype 1b clone in chimpanzees."

RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN C AND MRNA (BY SIMILARITY).

CC EMBL; AF333324; AAK08509.1; -

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; P00804; P00804.

DR PIR; P00329; P00329.

DR HSSP; P26663; 1NS3.

DR GO; GO:0016021; C:Integral to membrane; IEA.

DR GO; GO:0019028; C:Viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F:electon transporter activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003686; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006118; F:electron transport; IEA.

DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; F:transcription; IEA.

DR GO; GO:0019079; F:vital genome replication; IEA.

DR GO; GO:0019087; F:vital transformation; IEA.

DR InterPro; IPR009003; Cys_Ser_tryp_sln.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_NS5b.
 DR InterPro; IPR004109; Replicase_C29.
 DR InterPro; IPR007095; RNA_pol_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF00998; HCV_NS5a; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDc; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327007 KM; 053B9A53B0A335 CRC64;

Query Match 95.7%; Score 1902; DB 12; Length 3010;
 Best Local Similarity 94.2%; Pred. No. 2.2e-151;
 Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 2 LITLSPYKLLARLITWLOLITRVEHLOWVTPPLNVRGSDALITLTCVHPELFD 61
 |||||
 DB 828 LITLSPYKVLRLIWLQYFTRAEHLQWVPLNVRGSDALITLTCVHPELFD 887
 |||||
 QY 62 ITKLALAFGLMTVLAQGITVFFVRAQGLIRACMLVRKAGGHVQVNAFMKLAALTGT 121
 |||||
 DB 888 ITKLALALGLPLMTVLAQGITVFFVRAQGLIRACMLVRKAGGHVQVNAFMKLAALTGT 947
 |||||
 QY 122 VYVDLITPLQMAHAGLDLAVAVEPVTFSDMEVKIITWGADTAACDIIISGLPVASARRG 181
 |||||
 DB 948 YLYNHLTPLRDWAHAGLRDLAVAVEPVTFSDMEVKIITWGADTAACDIIISGLPVASARRG 1007
 |||||
 QY 182 REILGPAIDNEGQGRLLAPITAYSQQTRGLGCIITSLTGRDKNQESEVQVSTATQ 241
 |||||
 DB 1008 KEILGPAIDNEGQGRLLAPITAYSQQTRGLGCIITSLTGRDKNQESEVQVSTATQ 1067
 |||||
 QY 242 SPLATCVNGVCMVTHFGSGKTLGPKPFTQMTTNTDQDLVGMQAPPGARSMTPTCTCGS 301
 |||||
 DB 1068 SPLATCVNGVCMVTHFGSGKTLGPKPFTQMTTNTDQDLVGMQAPPGARSMTPTCTCGS 1127
 |||||
 QY 302 SDLYLVTRHADVIPIRRRGDSRGLSLSPRPVSYLKSGSSGGPILCPSSGHAAGIFRAAVCTR 361
 |||||
 DB 1128 SDLYLVTRHADVIPIRRRGDSRGLSLSPRPVSYLKSGSSGGPILCPSSGHAAGIFRAAVCTR 1187
 |||||
 QY 362 GVAKAVDFIPVESMETTMR 380
 |||||
 DB 1188 GVAKAVDFIPVESMETTMR 1206
 |||||

Search completed: May 6, 2004, 09:35:47
 Job time : 36.5407 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 14.914 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-12

Sequence: 1 ALLTSPYKVLARLIWML.....RGAKAVDFIPVESMETMR 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCITUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfillseq.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1887	95.0	2201	4	US-09-539-601-6
2	1887	95.0	2201	4	US-09-539-601-15
3	1887	95.0	3010	4	US-09-539-601-3
4	1887	95.0	3010	4	US-09-539-601-21
5	1887	95.0	3010	4	US-09-539-601-27
6	1881	94.7	1692	3	US-09-263-933-4
7	1881	94.7	1692	3	US-09-263-933-2
8	1881	94.7	2307	3	US-09-263-933-2
9	1881	94.7	2307	4	US-09-919-901-2
10	1880	94.6	3010	4	US-09-539-601-33
11	1878	94.5	1692	3	US-09-263-933-11
12	1878	94.5	1692	3	US-09-919-901-11
13	1878	94.5	2307	3	US-09-263-933-9
14	1878	94.5	2307	4	US-09-919-901-9
15	1869	94.1	1692	3	US-09-263-933-18
16	1869	94.1	1692	3	US-09-919-901-18
17	1869	94.1	2307	3	US-09-263-933-16
18	1869	94.1	2307	4	US-09-919-901-16
19	1869	94.1	3010	4	US-09-539-601-3
20	1823	91.7	2013	1	US-08-324-977-12
21	1823	91.7	2013	2	US-08-384-616-12
22	1823	91.7	2013	2	US-08-904-686A-12
23	1823	91.7	2013	3	US-09-315-850-12
24	1823	91.7	2201	4	US-08-952-981A-2
25	1823	91.7	2620	1	US-08-324-977-32
26	1823	91.7	2620	1	US-08-384-616-32
27	1823	91.7	2620	2	US-08-904-686A-32

28	1823	91.7	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1823	91.7	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1823	91.7	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1823	91.7	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1823	91.7	2621	3	US-08-315-850-36	Sequence 36, Appl
33	1823	91.7	3010	1	US-08-324-977-2	Sequence 2, Appl
34	1823	91.7	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1823	91.7	3010	2	US-08-384-616-2	Sequence 2, Appl
36	1823	91.7	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1823	91.7	3010	2	US-08-904-686A-2	Sequence 2, Appl
38	1823	91.7	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1823	91.7	3010	3	US-08-315-850-2	Sequence 2, Appl
40	1823	91.7	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1718	86.5	1648	1	US-08-188-281B-12	Sequence 12, Appl
42	1718	86.5	1648	5	PCT-US94-07280-12	Sequence 12, Appl
43	1718	86.5	1648	5	PCT-US95-01087-12	Sequence 12, Appl
44	1718	86.5	3011	1	US-08-188-281B-1	Sequence 1, Appl
45	1718	86.5	3011	1	US-08-453-552-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-539-601-6
; Sequence 6, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartschlagel, Ralf FM
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-539-601-6

Query Match 95.0%; Score 1887; DB 4; Length 2201;
Best Local Similarity 93.4%; Pred. No. 3.8e-181;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY	2	ALLTSPYKVLARLIWMLQYLIITRENAHQWITPLNVRGGRDAIILITCAVEBELIFD	61
DB	19	LLTTSFHYKFLARLIWMLQYFIITRAEHLQWITPLNVRGGRDAVILITCAIHELIFD	78
QY	62	ITKLLAIFGPIVMTQACITKVPYVRAQGLIRACMIVKAAAGHYQMAEMKLAALGT	121
DB	79	ITKLLAIFGPIVMTQACITKVPYVRAQGLIRACMIVKAAAGHYQMAEMKLAALGT	138
QY	122	YVYDILITPLQDAHAGLNDLAVAVEPVIFSDMEVKIITMGADTAACGDIIISGLPVASARG	181
DB	139	YVYDILITPLQDAHAGLNDLAVAVEPVIFSDMEVKIITMGADTAACGDIIISGLPVASARG	198
QY	182	REILIGPADNFEQGRRLIAPITANSQOTRGLIGITISLGRDNQVEGEVQVSTANTQ	241
DB	199	REIHIGPADNFEQGRRLIAPITANSQOTRGLIGITISLGRDNQVEGEVQVSTANTQ	258
QY	242	SLATCNQVGVCTVFGAGSKTLIAGPKPITQMTYTNVDQDLVGWQAPPGASMTPTCTG	301
DB	255	SLATCNQVGVCTVFGAGSKTLIAGPKPITQMTYTNVDQDLVGWQAPPGASMTPTCTG	318
QY	302	SDLYIVTRHADYIVRRRGRSGILSPRPVSYLKGSSGPIILCSGAVNGIFRAAVCTR	361
DB	319	SDLYIVTRHADYIVRRRGRSGILSPRPVSYLKGSSGPIILCSGAVNGIFRAAVCTR	378
QY	362	GVAKAVDFIPVESMETMR 380	

Db 379 GYAKAVDFVVESEMETTMR 397

RESULT 2

US-09-539-601-15
 ; Sequence 15, Application US/09539601C
 ; Patent No. 6630343
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartschlagel, Ralf FW
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 ; FILE REFERENCE: all sequences
 ; CURRENT APPLICATION NUMBER: US/09/539,601C
 ; EARLIER FILING DATE: 2001-08-30
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 15
 ; LENGTH: 2201
 ; TYPE: PRT
 ; ORGANISM: Hepatitis C virus
 ; US-09-539-601-15

Query Match

Best Local Similarity 95.0%; Score 1887; DB 4; Length 2201;
 Best Local Similarity 93.4%; Pred. No. 3,8e-181;
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLTSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGDAIILLTCAVHPELIFD 61
 DB 19 LTLTSPHYKFLARLIMWLQYFITRAEHLQVWIPPLNVRGGDAVILLTCAIHPELIFT 78
 QY 62 ITKLALIFGPMVLAQGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGT 121
 DB 79 ITKLALIFGPMVLAQGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGT 138
 QY 122 YVVDHLTPLODMNAGRLDAVAVEPVIFSDMEVKIITGADTAACGDIISGLPVASARG 181
 DB 139 YVVDHLTPLODMNAGRLDAVAVEPVIFSDMEVKIITGADTAACGDIISGLPVASARG 198
 QY 182 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTARG 241
 DB 199 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTARG 258
 QY 242 SFATCVNVCMTVYHAGSKTLAGPKGPIITQMTNVDDLVGMQAPGARSLTPTCTCGS 301
 DB 259 SFATCVNVCMTVYHAGSKTLAGPKGPIITQMTNVDDLVGMQAPGARSLTPTCTCGS 318
 QY 302 SDYLVTIRADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361
 DB 319 SDYLVTIRADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 378
 QY 362 GYAKAVDFVVESEMETTMR 380
 DB 379 GYAKAVDFVVESEMETTMR 397

RESULT 3

US-09-539-601-3
 ; Sequence 3, Application US/09539601C
 ; Patent No. 6630343
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartschlagel, Ralf FW
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 ; FILE REFERENCE: all sequences
 ; CURRENT APPLICATION NUMBER: US/09/539,601C
 ; EARLIER FILING DATE: 2001-08-30
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 3010
 ; TYPE: PRT

; ORGANISM: Hepatitis C virus
 ; US-09-539-601-3

Query Match

Best Local Similarity 95.0%; Score 1887; DB 4; Length 3010;
 Best Local Similarity 93.4%; Pred. No. 6.1e-181;
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLTSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGDAIILLTCAVHPELIFD 61
 DB 828 LTLTSPHYKFLARLIMWLQYFITRAEHLQVWIPPLNVRGGDAVILLTCAIHPELIFT 887
 QY 62 ITKLALIFGPMVLAQGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGT 121
 DB 888 ITKLALIFGPMVLAQGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGT 947
 QY 122 YVVDHLTPLODMNAGRLDAVAVEPVIFSDMEVKIITGADTAACGDIISGLPVASARG 181
 DB 948 YVVDHLTPLODMNAGRLDAVAVEPVIFSDMEVKIITGADTAACGDIISGLPVASARG 1007
 QY 182 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTARG 241
 DB 1008 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTARG 1067
 QY 242 SFATCVNVCMTVYHAGSKTLAGPKGPIITQMTNVDDLVGMQAPGARSLTPTCTCGS 301
 DB 1068 SFATCVNVCMTVYHAGSKTLAGPKGPIITQMTNVDDLVGMQAPGARSLTPTCTCGS 1127
 QY 302 SDYLVTIRADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361
 DB 1128 SDYLVTIRADYIPVRRGDSRGSLLSPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 1187
 QY 362 GYAKAVDFVVESEMETTMR 380
 DB 1188 GYAKAVDFVVESEMETTMR 1206

RESULT 4

US-09-539-601-21
 ; Sequence 21, Application US/09539601C
 ; Patent No. 6630343
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartschlagel, Ralf FW
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 ; FILE REFERENCE: all sequences
 ; CURRENT APPLICATION NUMBER: US/09/539,601C
 ; EARLIER FILING DATE: 2001-08-30
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 21
 ; LENGTH: 3010
 ; TYPE: PRT
 ; ORGANISM: Hepatitis C virus
 ; US-09-539-601-21

Query Match

Best Local Similarity 95.0%; Score 1887; DB 4; Length 3010;
 Best Local Similarity 93.4%; Pred. No. 6.1e-181;
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LTLTSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGDAIILLTCAVHPELIFD 61
 DB 828 LTLTSPHYKFLARLIMWLQYFITRAEHLQVWIPPLNVRGGDAVILLTCAIHPELIFT 887
 QY 62 ITKLALIFGPMVLAQGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGT 121
 DB 888 ITKLALIFGPMVLAQGITKVPYFVRAOGLIRACMLVRKAGHYVQMAFMKLAALTGT 947
 QY 122 YVVDHLTPLODMNAGRLDAVAVEPVIFSDMEVKIITGADTAACGDIISGLPVASARG 181
 DB 948 YVVDHLTPLODMNAGRLDAVAVEPVIFSDMEVKIITGADTAACGDIISGLPVASARG 1007
 QY 182 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKNQVEGEVQVSTARG 241

```
Db 1008 REIHGPRDSLEGGOWRLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 1067
Qy 242 SFLATCVNGVCTVYHGAAGSKTLAPKPGPITOMTYNNVDLVGQAPPGARSMTPTCGS 301
Db 1068 SFLATCVNGVCTVYHGAAGSKTLAPKPGPITOMTYNNVDLVGQAPPGARSMTPTCGS 1127
Qy 302 SDLYVTNRADVIYPRRRGDSRGSLSLSPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
Db 1128 SDLYVTNRADVIYPRRRGDSRGSLSLSPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 1187
Qy 362 GVAKAVDFIVESMETTMR 380
Db 1188 GVAKAVDFIVESMETTMR 1206
```

```
RESULT 5
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartsch, Ralf FM
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; EARLIER FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-27
```

```
Query Match 95.0%; Score 1887; DB 4; Length 3010;
Best Local Similarity 93.4%; Pred. No. 6.1e-181;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;
```

```
Qy 2 LITLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
Db 828 LITLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFD 887
Qy 62 ITKLILAFGLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLALGT 121
Db 888 ITKLILAFGLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLALGT 947
Qy 122 YVVDHLPLODMAGRLDAVAEVPFSDMEVKITMGADTAACGDIISGLPVSARG 181
Db 948 YVVDHLPLODMAGRLDAVAEVPFSDMEVKITMGADTAACGDIISGLPVSARG 1007
Qy 182 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241
Db 1008 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 1067
Qy 242 SFLATCVNGVCTVYHGAAGSKTLAPKPGPITOMTYNNVDLVGQAPPGARSMTPTCGS 301
Db 1068 SFLATCVNGVCTVYHGAAGSKTLAPKPGPITOMTYNNVDLVGQAPPGARSMTPTCGS 1127
Qy 302 SDLYVTNRADVIYPRRRGDSRGSLSLSPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
Db 1128 SDLYVTNRADVIYPRRRGDSRGSLSLSPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 1187
Qy 362 GVAKAVDFIVESMETTMR 380
Db 1188 GVAKAVDFIVESMETTMR 1206
```

```
RESULT 6
US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940
```

```
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-4
```

```
Query Match 94.7%; Score 1881; DB 3; Length 1692;
Best Local Similarity 93.1%; Pred. No. 1e-180;
Matches 353; Conservative 10; Mismatches 14; Indels 0; Gaps 0;
```

```
Qy 2 LITLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFD 61
Db 107 LITLSPYKVLARLIMWLQYLITRVEAHQVWIPPLNVRGGRDAIILLTCAVHPELIFD 166
Qy 62 ITKLILAFGLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLALGT 121
Db 167 ITKLILAFGLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLALGT 226
Qy 122 YVVDHLPLODMAGRLDAVAEVPFSDMEVKITMGADTAACGDIISGLPVSARG 181
Db 227 YVVDHLPLODMAGRLDAVAEVPFSDMEVKITMGADTAACGDIISGLPVSARG 286
Qy 182 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 241
Db 287 RETILGPADNFEQGWRLAPITAYSQOTRGLGCIITSLTGRDKNQVEGEVQVSTATQ 346
Qy 242 SFLATCVNGVCTVYHGAAGSKTLAPKPGPITOMTYNNVDLVGQAPPGARSMTPTCGS 301
Db 347 SFLATCVNGVCTVYHGAAGSKTLAPKPGPITOMTYNNVDLVGQAPPGARSMTPTCGS 406
Qy 302 SDLYVTNRADVIYPRRRGDSRGSLSLSPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 361
Db 407 SDLYVTNRADVIYPRRRGDSRGSLSLSPVSYLKSGSGGPLLCPGSHAVGIFRAAVCTR 466
Qy 362 GVAKAVDFIVESMETTMR 380
Db 467 GVAKAVDFIVESMETTMR 485
```

```
RESULT 7
US-09-919-901-4
; Sequence 4, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; EARLIER APPLICATION NUMBER: 09/263,933
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-4

Query Match      94.7%; Score 1881; DB 4; Length 1692;
Best Local Similarity 93.1%; Pred. No. 1e-180;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 61
DB LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 166
QY 62 ITKLAIIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALATGT 121
DB ITKLAIIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALATGT 226
QY 122 YVVDHLTPLODMWAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB YVVDHLTPLODMWAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 286
QY 182 REILGPADNFEQGRMLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVVSATQ 241
DB REILGPADNFEQGRMLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVVSATQ 346
QY 242 SFPLATCNGVCWTVFHGAGSKTLAGPKGPIITQWYTNVDQDLVGWQAPPGARSMTPTCTG 301
DB SFPLATCNGVCWTVFHGAGSKTLAGPKGPIITQWYTNVDQDLVGWQAPPGARSMTPTCTG 406
QY 302 SDIYVTRADVIYVRRRGRDSSGLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 361
DB SDIYVTRADVIYVRRRGRDSSGLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 466
QY 362 GVAKAVDFIPVESMETMTR 380
DB GVAKAVDFIPVESMETMTR 485

RESULT 8
US-09-263-933-2
; Sequence 2, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1999-03-08
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-2

Query Match      94.7%; Score 1881; DB 3; Length 2307;
Best Local Similarity 93.1%; Pred. No. 1.7e-180;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 61
DB LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 256
QY 199 LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 256
DB LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 318
QY 62 ITKLAIIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALATGT 121
DB ITKLAIIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALATGT 318
```

```

QY 122 YVVDHLTPLODMWAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB YVVDHLTPLODMWAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPADNFEQGRMLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVVSATQ 241
DB REILGPADNFEQGRMLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVVSATQ 438
QY 242 SFPLATCNGVCWTVFHGAGSKTLAGPKGPIITQWYTNVDQDLVGWQAPPGARSMTPTCTG 301
DB SFPLATCNGVCWTVFHGAGSKTLAGPKGPIITQWYTNVDQDLVGWQAPPGARSMTPTCTG 498
QY 302 SDIYVTRADVIYVRRRGRDSSGLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 361
DB SDIYVTRADVIYVRRRGRDSSGLSPRPVSYLKSSGGPILCPSGHAGVIFRAAVCTR 558
QY 362 GVAKAVDFIPVESMETMTR 380
DB GVAKAVDFIPVESMETMTR 577

RESULT 9
US-09-919-901-2
; Sequence 2, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; EARLIER FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-2

Query Match      94.7%; Score 1881; DB 4; Length 2307;
Best Local Similarity 93.1%; Pred. No. 1.7e-180;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 61
DB LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 258
QY 199 LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 258
DB LTLSPYKVLARLIMLQYITRVEAHLQVWIPPLNVRGGRDAIILTCVAHPBLIFD 318
QY 62 ITKLAIIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALATGT 121
DB ITKLAIIFGPIPLMVLQAGITKVPYFRAOGLIRACMLYKKAAGHYVQMAFKALATGT 318
QY 122 YVVDHLTPLODMWAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB YVVDHLTPLODMWAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPADNFEQGRMLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVVSATQ 241
DB REILGPADNFEQGRMLAPITAYSQOTRGLGCIITSLTGRDKNOVEGEVQVVSATQ 438
QY 242 SFPLATCNGVCWTVFHGAGSKTLAGPKGPIITQWYTNVDQDLVGWQAPPGARSMTPTCTG 301
DB SFPLATCNGVCWTVFHGAGSKTLAGPKGPIITQWYTNVDQDLVGWQAPPGARSMTPTCTG 498
```


QY 302 SDLVYVTHADVIYVRRRGRSGSLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361
 |||||
 DB 499 SDLVYVTHADVIYVRRRGRSGSLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 558
 |||||
 QY 362 GVAKAVDFVPEVSMETMR 380
 |||||
 DB 559 GVAKAVDFVPEVSMETMR 577
 |||||

RESULT 10
 US-09-539-601-33

/ Sequence 33, Application US/09539601C
 / Patent No. 6630343
 / GENERAL INFORMATION:
 / APPLICANT: Bartschlager, Ralf FW
 / TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 / FILE REFERENCE: all sequences
 / CURRENT APPLICATION NUMBER: US/09/539,601C
 / CURRENT FILING DATE: 2001-08-30
 / EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 / EARLIER FILING DATE: 1999-04-03
 / NUMBER OF SEQ ID NOS: 51
 / SOFTWARE: Patentin Ver. 2.1
 / SEQ ID NO 33
 / LENGTH: 3010
 / TYPE: PRT
 / ORGANISM: Hepatitis C virus
 US-09-539-601-33

Query Match 94.6%; Score 1880; DB 4; Length 3010;
 Best Local Similarity 93.1%; Pred. No. 3.1e-180;

Matches 353; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMWLOYLITRVEAHQWIPPLNVRGGRDAIILTCVAHPELIFD 61
 |||||
 DB 828 LTLSPYKVLARLIMWLOYLITRVEAHQWIPPLNVRGGRDAIILTCVAHPELIFD 887
 |||||
 QY 62 ITKLILAFGRLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLAALTGT 121
 |||||
 DB 888 ITKLILAFGRLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLAALTGT 947
 |||||
 QY 122 YVVDHLPLQDMARAGRLDAVAEVPVPSDMEVKIITMGADTAACGDIISGLPVSARG 181
 |||||
 DB 948 YVVDHLPLQDMARAGRLDAVAEVPVPSDMEVKIITMGADTAACGDIISGLPVSARG 1007
 |||||
 QY 182 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKQVGEVQVSTATQ 241
 |||||
 DB 1008 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKQVGEVQVSTATQ 1067
 |||||
 QY 242 SFLATCVNGVCMVYFHGAGSKITLAGPKGPIITOMYTNVDOLVGMQAPGARSLPCTCGS 301
 |||||
 DB 1068 SFLATCVNGVCMVYFHGAGSKITLAGPKGPIITOMYTNVDOLVGMQAPGARSLPCTCGS 1127
 |||||
 QY 302 SDLVYVTHADVIYVRRRGRSGSLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361
 |||||
 DB 1128 SDLVYVTHADVIYVRRRGRSGSLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 1187
 |||||
 QY 362 GVAKAVDFVPEVSMETMR 380
 |||||
 DB 1188 GVAKAVDFVPEVSMETMR 1206
 |||||

RESULT 11
 US-09-263-933-11

/ Sequence 11, Application US/09263933
 / Patent No. 6280940
 / GENERAL INFORMATION:
 / APPLICANT: Potts, Karen E.
 / APPLICANT: Patrick, Roberta L.
 / TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 / TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 / FILE REFERENCE: 0125-0005A

/ CURRENT APPLICATION NUMBER: US/09/263,933
 / CURRENT FILING DATE: 1999-03-08
 / EARLIER APPLICATION NUMBER: 09/129,611
 / EARLIER FILING DATE: 1998-08-05
 / NUMBER OF SEQ ID NOS: 33
 / SOFTWARE: Patentin Ver. 2.0
 / SEQ ID NO 11
 / LENGTH: 1692
 / TYPE: PRT
 / ORGANISM: Artificial Sequence
 US-09-263-933-11

Query Match 94.5%; Score 1878; DB 3; Length 1692;
 Best Local Similarity 92.9%; Pred. No. 2.1e-180;
 Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

QY 2 LTLSPYKVLARLIMWLOYLITRVEAHQWIPPLNVRGGRDAIILTCVAHPELIFD 61
 |||||
 DB 107 LTLSPYKVLARLIMWLOYLITRVEAHQWIPPLNVRGGRDAIILTCVAHPELIFD 166
 |||||
 QY 62 ITKLILAFGRLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLAALTGT 121
 |||||
 DB 167 ITKLILAFGRLMVLQAGITKVPYFVRAOGLIRACMLVRKAAGHYVQAFMKLAALTGT 226
 |||||
 QY 122 YVVDHLPLQDMARAGRLDAVAEVPVPSDMEVKIITMGADTAACGDIISGLPVSARG 181
 |||||
 DB 227 YVVDHLPLQDMARAGRLDAVAEVPVPSDMEVKIITMGADTAACGDIISGLPVSARG 286
 |||||
 QY 182 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKQVGEVQVSTATQ 241
 |||||
 DB 287 REILGPADNFEQGWRLAPITAYSOQTRGLGCIITSLTGRDKQVGEVQVSTATQ 346
 |||||
 QY 242 SFLATCVNGVCMVYFHGAGSKITLAGPKGPIITOMYTNVDOLVGMQAPGARSLPCTCGS 301
 |||||
 DB 347 SFLATCVNGVCMVYFHGAGSKITLAGPKGPIITOMYTNVDOLVGMQAPGARSLPCTCGS 406
 |||||
 QY 302 SDLVYVTHADVIYVRRRGRSGSLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 361
 |||||
 DB 407 SDLVYVTHADVIYVRRRGRSGSLSPRPVSYLKSGSGPILCPGSHAVGIFRAAVCTR 466
 |||||
 QY 362 GVAKAVDFVPEVSMETMR 380
 |||||
 DB 467 GVAKAVDFVPEVSMETMR 485
 |||||

RESULT 12

US-09-919-901-11

/ Sequence 11, Application US/09919901
 / Patent No. 6599738
 / GENERAL INFORMATION:
 / APPLICANT: Potts, Karen E.
 / APPLICANT: Patrick, Roberta L.
 / TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 / TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 / FILE REFERENCE: 0125-0005A
 / CURRENT APPLICATION NUMBER: US/09/919,901
 / CURRENT FILING DATE: 2001-08-02
 / PRIOR APPLICATION NUMBER: 09/263,933
 / PRIOR FILING DATE: 1999-02-08
 / PRIOR APPLICATION NUMBER: 09/129,611
 / PRIOR FILING DATE: 1998-08-05
 / NUMBER OF SEQ ID NOS: 33
 / SOFTWARE: Patentin Ver. 2.0
 / SEQ ID NO 11
 / LENGTH: 1692
 / TYPE: PRT
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: :
 US-09-919-901-11

Query Match 94.5%; Score 1878; DB 4; Length 1692;

Best Local Similarity 92.9%; Pred. No. 2,1e-180;
Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

```

QY 2 LTLSPYKVLARLIMWLYLITRVEAHLQVWIPPLNVRGGSDAIILLTCVHPELIFD 61
DB 107 LTLSPYKVLARLIMWLYLITRVEAHLQVWIPPLNVRGGSDAIILLTCVHPELIFD 166
QY 62 ITKLLAIFGPLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
DB 167 ITKLLAIFGPLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 226
QY 122 YVVDHLTFLQDMWHAAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB 227 YVVDHLTFLQDMWHAAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 286
QY 182 REILGPDNFEQGGWRLAPITAVSQOTRGLGCIITSLTGRDKNQEVEGVVSTATQ 241
DB 287 KEILGPDNFEQGGWRLAPITAVSQOTRGLGCIITSLTGRDKNQEVEGVVSTATQ 346
QY 242 SFPLATCVGCVMTVYHAGSKTLGAGKGPITQMTYTNVDDLVGMQAPPGARSLTPTCGS 301
DB 347 SFPLATCVGCVMTVYHAGSKTLGAGKGPITQMTYTNVDDLVGMQAPPGARSLTPTCGS 406
QY 302 SDLYLVRHADVIIVRRRDSRGLSLPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 407 SDLYLVRHADVIIVRRRDSRGLSLPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 466
QY 362 GVAKAVDFIVPESMETTMR 380
DB 467 GVAKAVDFIVPESMETTMR 485

```

RESULT 13

US-09-263-933-9
Sequence 9, Application US/09263933

```

; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-9

```

Query Match 94.5%; Score 1878; DB 3; Length 2307;
Best Local Similarity 92.9%; Pred. No. 3,3e-180;
Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

```

QY 2 LTLSPYKVLARLIMWLYLITRVEAHLQVWIPPLNVRGGSDAIILLTCVHPELIFD 61
DB 199 LTLSPYKVLARLIMWLYLITRVEAHLQVWIPPLNVRGGSDAIILLTCVHPELIFD 258
QY 62 ITKLLAIFGPLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
DB 259 ITKLLAIFGPLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 318
QY 122 YVVDHLTFLQDMWHAAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB 319 YVVDHLTFLQDMWHAAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPDNFEQGGWRLAPITAVSQOTRGLGCIITSLTGRDKNQEVEGVVSTATQ 241
DB 577 REILGPDNFEQGGWRLAPITAVSQOTRGLGCIITSLTGRDKNQEVEGVVSTATQ 577

```

```

DB 379 KEILGPDNFEQGGWRLAPITAVSQOTRGLGCIITSLTGRDKNQEVEGVVSTATQ 438
QY 242 SFPLATCVGCVMTVYHAGSKTLGAGKGPITQMTYTNVDDLVGMQAPPGARSLTPTCGS 301
DB 439 SFPLATCVGCVMTVYHAGSKTLGAGKGPITQMTYTNVDDLVGMQAPPGARSLTPTCGS 498
QY 302 SDLYLVRHADVIIVRRRDSRGLSLPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 499 SDLYLVRHADVIIVRRRDSRGLSLPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 558
QY 362 GVAKAVDFIVPESMETTMR 380
DB 559 GVAKAVDFIVPESMETTMR 577

```

RESULT 14

US-09-919-901-9
Sequence 9, Application US/09919901

```

; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-9

```

Query Match 94.5%; Score 1878; DB 4; Length 2307;
Best Local Similarity 92.9%; Pred. No. 3,3e-180;
Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

```

QY 2 LTLSPYKVLARLIMWLYLITRVEAHLQVWIPPLNVRGGSDAIILLTCVHPELIFD 61
DB 199 LTLSPYKVLARLIMWLYLITRVEAHLQVWIPPLNVRGGSDAIILLTCVHPELIFD 258
QY 62 ITKLLAIFGPLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 121
DB 259 ITKLLAIFGPLMVLQAGITRVYFVRAQGLIRACMLVRKAAGHYVQMAFMKLAALTGT 318
QY 122 YVVDHLTFLQDMWHAAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 181
DB 319 YVVDHLTFLQDMWHAAGRLDAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVASARG 378
QY 182 REILGPDNFEQGGWRLAPITAVSQOTRGLGCIITSLTGRDKNQEVEGVVSTATQ 241
DB 379 REILGPDNFEQGGWRLAPITAVSQOTRGLGCIITSLTGRDKNQEVEGVVSTATQ 438
QY 242 SFPLATCVGCVMTVYHAGSKTLGAGKGPITQMTYTNVDDLVGMQAPPGARSLTPTCGS 301
DB 439 SFPLATCVGCVMTVYHAGSKTLGAGKGPITQMTYTNVDDLVGMQAPPGARSLTPTCGS 498
QY 302 SDLYLVRHADVIIVRRRDSRGLSLPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 499 SDLYLVRHADVIIVRRRDSRGLSLPRPVSYLKSSGGPILCPGSHAVGIFRAAVCTR 558
QY 362 GVAKAVDFIVPESMETTMR 380
DB 559 GVAKAVDFIVPESMETTMR 577

```

RESULT 15

US-09-263-933-18
Sequence 18, Application US/09263933
Patent No. 6280940
GENERAL INFORMATION:
APPLICANT: Echts, Karen E.
APPLICANT: Jackson, Rodetta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/263,933
CURRENT FILING DATE: 1999-03-08
EARLIER APPLICATION NUMBER: 09/129,611
EARLIER FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO: 18
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
US-09-263-933-18

Query Match 94.1%; Score 1869; DB 3; Length 1692;

Best Local Similarity 92.6%; Pred. No. 1,7e-179;

Matches 351; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 2 LTTLSPPYVYVLLARLIMWLQYLITRVEAHLQVWIPPLNVRGGDAIILITCAVHPELIID 61
DB 107 LTTLSPPYVYVFLARLIMWLQYFTTRAEALHWIPPLNVRGGDAIILITCAVHPELIID 166
QY 62 ITKLLAIIGPLIMVQAGITKPYFPAQGLRACMLVPRKAGHYVQMAFYKLAALTGT 121
DB 167 ITKLLAIIGPLIMVQAGITRVPYFPAQGLIACMLVPRKAGHYVQMAFYKLAALTGT 226
QY 122 YVVDHLTPQDWAHAGLRDLAVALVEPVIFSDMEVKIITWADTAAAGDIISGLPVARSRG 181
DB 227 YVVDHLTPQDWAHAGLRDLAVALVEPVIFSDMEVKIITWADTAAAGDIISGLPVARSRG 286
QY 182 REILGPDNFBGQWRLAPITAYSQTRGLGCIITSLTGRDKQVGEVQVYSTATQ 241
DB 287 KEILGPDNFBGQWRLAPITAYSQTRGLGCIITSLTGRDKQVGEVQVYSTATQ 346
QY 242 SFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDDLVGMQAPPGARSMTPTCTGSS 301
DB 347 SFATCVNGVCMVTFHAGSKTLAGPKGPIITQMTNVDDLVGMQAPPGARSMTPTCTGSS 406
QY 302 SDLYVTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCPGHAAGIFRAVCTR 361
DB 407 SDLYVTRHADVI PVRRGDSRGLSPRPVSYLKSSGGPILCPGHAAGIFRAVCTR 466
QY 362 GVAKAVDFIVBSMETTMR 380
DB 467 GVAKAVDFIVBSMETTMR 485

Search completed: May 6, 2004, 09:39:03
Job time: 15.9144 secs

Fri May 7 13:37:02 2004

us-10-650-585-12.rapb

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 38.8727 Seconds
(without alignments)
2713.357 Million cell updates/sec

Title: US-10-650-585-12
Perfect score: 1987
Sequence: 1 ALLTSPYKVLARLIMWL.....RGVAKAVDFIVESNETTR 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 27566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1987	100.0	380	US-10-017-736-12	Sequence 12, Appl
2	1987	100.0	380	US-10-650-585-12	Sequence 12, Appl
3	1987	100.0	393	US-10-017-736-11	Sequence 11, Appl
4	1987	100.0	393	US-10-650-585-11	Sequence 11, Appl
5	1987	100.0	409	US-10-017-736-2	Sequence 2, Appl
6	1987	100.0	409	US-10-650-585-2	Sequence 2, Appl
7	1887	95.0	2201	US-10-029-907-3	Sequence 3, Appl
8	1887	95.0	2201	US-10-309-561-3	Sequence 3, Appl
9	1887	95.0	3010	US-10-467-000-1	Sequence 1, Appl
10	1881	94.7	1692	US-09-919-901-4	Sequence 4, Appl
11	1881	94.7	1692	US-10-191-966-1	Sequence 4, Appl
12	1881	94.7	2307	US-09-919-901-2	Sequence 2, Appl
13	1881	94.7	2307	US-10-191-966-2	Sequence 2, Appl
14	1878	94.5	1692	US-09-919-901-11	Sequence 11, Appl
15	1878	94.5	1692	US-10-191-966-11	Sequence 11, Appl

16	1878	94.5	2307	US-09-919-901-9	Sequence 9, Appl
17	1878	94.5	2307	US-10-191-966-9	Sequence 9, Appl
18	1869	94.1	1692	US-09-919-901-18	Sequence 18, Appl
19	1869	94.1	1692	US-10-191-966-18	Sequence 18, Appl
20	1869	94.1	2307	US-09-919-901-16	Sequence 16, Appl
21	1869	94.1	2307	US-10-191-966-16	Sequence 16, Appl
22	1842	92.7	352	US-10-017-736-13	Sequence 13, Appl
23	1842	92.7	352	US-10-650-585-13	Sequence 13, Appl
24	1823	91.7	2201	US-10-085-476-2	Sequence 2, Appl
25	1823	91.7	2201	US-10-017-736-14	Sequence 14, Appl
26	1778	89.5	341	US-10-650-585-14	Sequence 4, Appl
27	1717	86.4	3011	US-09-742-659-4	Sequence 4, Appl
28	1717	86.4	3011	US-09-891-894-3	Sequence 3, Appl
29	1717	86.4	3011	US-10-184-150-3	Sequence 3, Appl
30	1717	86.4	3011	US-10-328-997-3	Sequence 3, Appl
31	1717	86.4	3012	US-08-238-076-2	Sequence 2, Appl
32	1717	86.4	3012	US-09-895-937-2	Sequence 2, Appl
33	1717	86.4	3012	US-09-917-563-2	Sequence 2, Appl
34	1715	86.3	3011	US-09-916-359-2	Sequence 2, Appl
35	1715	86.3	3011	US-10-296-734-406	Sequence 406, App
36	1713	86.2	3011	US-09-238-076-20	Sequence 20, Appl
37	1713	86.2	3011	US-09-895-937-20	Sequence 20, Appl
38	1713	86.2	3011	US-09-917-563-20	Sequence 20, Appl
39	1710	86.1	2894	US-09-941-611-23	Sequence 23, Appl
40	1710	86.1	2894	US-10-044-995-23	Sequence 23, Appl
41	1710	86.1	3011	US-09-952-572-9	Sequence 9, Appl
42	1710	86.1	3011	US-09-747-419-20	Sequence 20, Appl
43	1710	86.1	3011	US-10-189-359-14	Sequence 14, Appl
44	1710	86.1	3011	US-10-259-275-20	Sequence 20, Appl
45	1707	85.9	2985	US-10-259-275-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-12

Query Match	100.0%	Score 1987	DB 13	Length 380
Best Local Similarity	100.0%	Pred. No. 1.2e+13	Indels 0	Gaps 0
Matches 380	Conservative	0	Mismatches 0	
QY	1	ALLTSPYKVLARLIMWLQYLITREVAHLQWIPINVRGGRDAIILLTCVAPPELIF	60	
DB	1	ALLTSPYKVLARLIMWLQYLITREVAHLQWIPINVRGGRDAIILLTCVAPPELIF	60	
QY	61	DIITKLALFGLNVLQAGITVPYFVRAOGLIRACMLYRKAAGHYQMAFKLAALTG	120	
DB	61	DIITKLALFGLNVLQAGITVPYFVRAOGLIRACMLYRKAAGHYQMAFKLAALTG	120	
QY	121	TYVYDHLTFLQWMAHAGLDLVAVEPVFSPMEVKIITWAGDTAACGIIISGLPVSARR	180	
DB	121	TYVYDHLTFLQWMAHAGLDLVAVEPVFSPMEVKIITWAGDTAACGIIISGLPVSARR	180	
QY	181	GREILGPADNEGGQWRLLAPITAVSQTRGLGCIITSLTGRDNQVGEVQVSTNT	240	
DB	181	GREILGPADNEGGQWRLLAPITAVSQTRGLGCIITSLTGRDNQVGEVQVSTNT	240	

Db 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 240
 QY 241 QSFATCVNGCVMTVFHAGSKTLGPKGPIOTMTNVDDLVGMQAPPGARSMTPTCTG 300
 Db 241 QSFATCVNGCVMTVFHAGSKTLGPKGPIOTMTNVDDLVGMQAPPGARSMTPTCTG 300
 QY 301 SSDLYVTRHADYIPVRRRDSRGLSPRPVSYLKGSSGGLPCPSGHAAGIFRAAVCT 360
 Db 301 SSDLYVTRHADYIPVRRRDSRGLSPRPVSYLKGSSGGLPCPSGHAAGIFRAAVCT 360
 QY 361 RGVAKAVDFIPVESMETMR 380
 Db 361 RGVAKAVDFIPVESMETMR 380

RESULT 2

US-10-650-585-12
 ; Sequence 12, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 380
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-650-585-12

Query Match 100.0%; Score 1987; DB 16; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1,2e-193;
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIIF 60
 Db 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIIF 60
 QY 61 DITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVRKKAAGHYQMAFMKLAALTG 120
 Db 61 DITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVRKKAAGHYQMAFMKLAALTG 120
 QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180
 Db 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180
 QY 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 240
 Db 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 240
 QY 241 QSFATCVNGCVMTVFHAGSKTLGPKGPIOTMTNVDDLVGMQAPPGARSMTPTCTG 300
 Db 241 QSFATCVNGCVMTVFHAGSKTLGPKGPIOTMTNVDDLVGMQAPPGARSMTPTCTG 300
 QY 301 SSDLYVTRHADYIPVRRRDSRGLSPRPVSYLKGSSGGLPCPSGHAAGIFRAAVCT 360
 Db 301 SSDLYVTRHADYIPVRRRDSRGLSPRPVSYLKGSSGGLPCPSGHAAGIFRAAVCT 360
 QY 361 RGVAKAVDFIPVESMETMR 380
 Db 361 RGVAKAVDFIPVESMETMR 380

RESULT 3

US-10-017-736-11
 ; Sequence 11, Application US/10017736

; Publication No. US20020192640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/017,736
 ; CURRENT FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 393
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-017-736-11

Query Match 100.0%; Score 1987; DB 13; Length 393;
 Best Local Similarity 100.0%; Pred. No. 1,3e-193;
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIIF 60
 Db 1 ALLTSPYKVLARLIWMLQYLITRVEAHLOVWIPPLNVRGGRDAIILLTCVHPELIIF 73
 QY 61 DITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVRKKAAGHYQMAFMKLAALTG 120
 Db 61 DITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLVRKKAAGHYQMAFMKLAALTG 133
 QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 180
 Db 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSARR 193
 QY 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 240
 Db 181 GREILLGPADNFEQGGKRLAPITAYSQOTRGLGCIITSLTRDKNOVEGEVQVSTAT 253
 QY 241 QSFATCVNGCVMTVFHAGSKTLGPKGPIOTMTNVDDLVGMQAPPGARSMTPTCTG 300
 Db 241 QSFATCVNGCVMTVFHAGSKTLGPKGPIOTMTNVDDLVGMQAPPGARSMTPTCTG 313
 QY 301 SSDLYVTRHADYIPVRRRDSRGLSPRPVSYLKGSSGGLPCPSGHAAGIFRAAVCT 360
 Db 301 SSDLYVTRHADYIPVRRRDSRGLSPRPVSYLKGSSGGLPCPSGHAAGIFRAAVCT 373
 QY 361 RGVAKAVDFIPVESMETMR 380
 Db 361 RGVAKAVDFIPVESMETMR 393

RESULT 4

US-10-650-585-11
 ; Sequence 11, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 393
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-650-585-11

Query Match 100.0%; Score 1987; DB 16; Length 393;

Fri May 7 13:37:02 2004

us-10-650-585-12.rapb

Page 3

Best Local Similarity 100.0%; Pred. No. 1.3e-193;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60
Db 14 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 73
QY 61 DITLLAIIGPLMVLQAGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120
Db 74 DITLLAIIGPLMVLQAGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 133
QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAR 180
Db 134 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAR 193
QY 181 GREILLGPADNFEQGRRLAPITAYSOQTRGLGCIITSLGRDNQVGEVQVSTAT 240
Db 194 GREILLGPADNFEQGRRLAPITAYSOQTRGLGCIITSLGRDNQVGEVQVSTAT 253
QY 241 QSFILATCNGVCWTVFPGAGSKTLAAGPKGITOMYTNVDDLVGMQAPPGARSMTPTCTG 300
Db 254 QSFILATCNGVCWTVFPGAGSKTLAAGPKGITOMYTNVDDLVGMQAPPGARSMTPTCTG 313
QY 301 SSDIYLVTRADVIPIVRRGDSRGSLSPPVSYLKSGSGGPLLCPGSHAVGIFRAAVCT 360
Db 314 SSDIYLVTRADVIPIVRRGDSRGSLSPPVSYLKSGSGGPLLCPGSHAVGIFRAAVCT 373
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 374 RGVAKAVDFIPVESMETTMR 393
```

RESULT 5
US-10-017-736-2
Sequence 2, Application US/10017736
Publication No. US20020192640A1
GENERAL INFORMATION:
APPLICANT: Boehringer Ingelheim (Canada) Ltd.
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
FILE REFERENCE: 13/082
CURRENT APPLICATION NUMBER: US/10/017,736
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: 60/256,031
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 409
TYPE: PRT
ORGANISM: HCV
US-10-017-736-2

Query Match 100.0%; Score 1987; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 1.4e-193;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60
Db 18 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 77
QY 61 DITLLAIIGPLMVLQAGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120
Db 78 DITLLAIIGPLMVLQAGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 137
QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAR 180
Db 138 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAR 197
QY 181 GREILLGPADNFEQGRRLAPITAYSOQTRGLGCIITSLGRDNQVGEVQVSTAT 240
Db 198 GREILLGPADNFEQGRRLAPITAYSOQTRGLGCIITSLGRDNQVGEVQVSTAT 257
QY 241 QSFILATCNGVCWTVFPGAGSKTLAAGPKGITOMYTNVDDLVGMQAPPGARSMTPTCTG 300
```

Best Local Similarity 100.0%; Pred. No. 1.4e-193;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 258 QSFILATCNGVCWTVFPGAGSKTLAAGPKGITOMYTNVDDLVGMQAPPGARSMTPTCTG 317
QY 301 SSDIYLVTRADVIPIVRRGDSRGSLSPPVSYLKSGSGGPLLCPGSHAVGIFRAAVCT 360
Db 318 SSDIYLVTRADVIPIVRRGDSRGSLSPPVSYLKSGSGGPLLCPGSHAVGIFRAAVCT 377
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 378 RGVAKAVDFIPVESMETTMR 397
```

RESULT 6
US-10-650-585-2
Sequence 2, Application US/10650585
Publication No. US20040077066A1
GENERAL INFORMATION:
APPLICANT: Boehringer Ingelheim (Canada) Ltd.
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
FILE REFERENCE: 13/082
CURRENT APPLICATION NUMBER: US/10/650,585
PRIOR FILING DATE: 2003-08-28
PRIOR APPLICATION NUMBER: US/10/017,736A
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: 60/256,031
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 409
TYPE: PRT
ORGANISM: HCV
US-10-650-585-2

Query Match 100.0%; Score 1987; DB 16; Length 409;
Best Local Similarity 100.0%; Pred. No. 1.4e-193;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 60
Db 18 ALTLSPYKVLARLIMWLQYLITRVEAHLQVWIPPLNVRGGRDAIILLTCAVHPELIF 77
QY 61 DITLLAIIGPLMVLQAGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 120
Db 78 DITLLAIIGPLMVLQAGITKPYFVRAQGLIRACMLVRKKAAGHYVQMAFMKLAALTG 137
QY 121 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAR 180
Db 138 TYVVDHLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITWGADTAACGDIISGLPVSAR 197
QY 181 GREILLGPADNFEQGRRLAPITAYSOQTRGLGCIITSLGRDNQVGEVQVSTAT 240
Db 198 GREILLGPADNFEQGRRLAPITAYSOQTRGLGCIITSLGRDNQVGEVQVSTAT 257
QY 241 QSFILATCNGVCWTVFPGAGSKTLAAGPKGITOMYTNVDDLVGMQAPPGARSMTPTCTG 300
Db 258 QSFILATCNGVCWTVFPGAGSKTLAAGPKGITOMYTNVDDLVGMQAPPGARSMTPTCTG 317
QY 301 SSDIYLVTRADVIPIVRRGDSRGSLSPPVSYLKSGSGGPLLCPGSHAVGIFRAAVCT 360
Db 318 SSDIYLVTRADVIPIVRRGDSRGSLSPPVSYLKSGSGGPLLCPGSHAVGIFRAAVCT 377
QY 361 RGVAKAVDFIPVESMETTMR 380
Db 378 RGVAKAVDFIPVESMETTMR 397
```

RESULT 7
US-10-029-907-3
Sequence 3, Application US/10029907
Publication No. US20020142350A1
GENERAL INFORMATION:
APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.


```

/ TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
/ FILE REFERENCE: 13/083
/ CURRENT APPLICATION NUMBER: US/10/029,907
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,857
/ PRIOR FILING DATE: 2000-12-22
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: FaSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 2201
/ TYPE: PRT
/ ORGANISM: HCV
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: 882
/ OTHER INFORMATION: Xaa is Lys or Arg
/ NAME/KEY: VARIANT
/ LOCATION: 1489
/ OTHER INFORMATION: Xaa is Leu
US-10-029-907-3

```

```

Query Match          95.0%; Score 1887; DB 13; Length 2201;
Best Local Similarity 93.4%; Pred. No. 2,3e-182;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```

```

QY 2 LITLSPYKVLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFD 61
DB 19 LITLSPYKVLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFD 78
QY 62 ITKLAIIFGPIWMLQGITKVPYFVRAAGLIRACMLVRKAGHVVQMAFMKLAALTGT 121
DB 79 ITKLAIIFGPIWMLQGITKVPYFVRAAGLIRACMLVRKAGHVVQMAFMKLAALTGT 138
QY 122 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 181
DB 139 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 198
QY 182 REILGPADNPEGQWRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 241
DB 199 REILGPADNPEGQWRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 258
QY 242 SFATCVNGVCMVTFVHAGSKTLAGPKGPIITQMTYTNVDQDLVGMQAPPGARSYTPCTCGS 301
DB 259 SFATCVNGVCMVTFVHAGSKTLAGPKGPIITQMTYTNVDQDLVGMQAPPGARSYTPCTCGS 318
QY 302 SDLYLTVRHADVIPIVRRGDSRGLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 319 SDLYLTVRHADVIPIVRRGDSRGLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 378
QY 362 GVAKAVDFIVPESMETTR 380
DB 379 GVAKAVDFIVPESMETTR 397

```

```

RESULT 8
US-10-309-561-3
/ Sequence 3, Application US/10309561
/ Publication No. US20030148348A1
/ GENERAL INFORMATION:
/ APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
/ TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
/ FILE REFERENCE: 13/083
/ CURRENT APPLICATION NUMBER: US/10/309,561
/ CURRENT FILING DATE: 2002-12-04
/ PRIOR APPLICATION NUMBER: US/10/029,907
/ PRIOR FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,857
/ PRIOR FILING DATE: 2000-12-22
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: FaSeq for Windows Version 4.0
/ SEQ ID NO 3

```

```

/ LENGTH: 2201
/ TYPE: PRT
/ ORGANISM: HCV
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: 882
/ OTHER INFORMATION: Xaa is Lys or Arg
/ NAME/KEY: VARIANT
/ LOCATION: 1489
/ OTHER INFORMATION: Xaa is Leu
US-10-309-561-3

```

```

Query Match          95.0%; Score 1887; DB 14; Length 2201;
Best Local Similarity 93.4%; Pred. No. 2,3e-182;
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

```

```

QY 2 LITLSPYKVLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFD 61
DB 19 LITLSPYKVLARLWMLQYITRVAHLQWIPPLNVRGGRDAIILTCVHPELIFD 78
QY 62 ITKLAIIFGPIWMLQGITKVPYFVRAAGLIRACMLVRKAGHVVQMAFMKLAALTGT 121
DB 79 ITKLAIIFGPIWMLQGITKVPYFVRAAGLIRACMLVRKAGHVVQMAFMKLAALTGT 138
QY 122 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 181
DB 139 YVVDHLTPLODMAHAGRLDAVAVEPVFSDEMEVKITWGADTAAACGDIILGLPVASARG 198
QY 182 REILGPADNPEGQWRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 241
DB 199 REILGPADNPEGQWRLAPITAYSOQTRGLGCIITSITGRDNQVEGEVGVSTATQ 258
QY 242 SFATCVNGVCMVTFVHAGSKTLAGPKGPIITQMTYTNVDQDLVGMQAPPGARSYTPCTCGS 301
DB 259 SFATCVNGVCMVTFVHAGSKTLAGPKGPIITQMTYTNVDQDLVGMQAPPGARSYTPCTCGS 318
QY 302 SDLYLTVRHADVIPIVRRGDSRGLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 361
DB 319 SDLYLTVRHADVIPIVRRGDSRGLSPRVSYLTKSSGGPILCPGSHAVGIFRAAVCTR 378
QY 362 GVAKAVDFIVPESMETTR 380
DB 379 GVAKAVDFIVPESMETTR 397

```

```

RESULT 9
US-10-467-000-1
/ Sequence 1, Application US/10467000
/ Publication No. US20040067486A1
/ GENERAL INFORMATION:
/ APPLICANT: De Francesco, Raffaele
/ APPLICANT: Migliaccio, Giovanni
/ APPLICANT: Paonessa, Giacomo
/ TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
/ FILE REFERENCE: ITR0003P
/ CURRENT APPLICATION NUMBER: US/10/467,000
/ CURRENT FILING DATE: 2003-07-21
/ PRIOR APPLICATION NUMBER: PCT/EP02/00526
/ PRIOR FILING DATE: 2002-01-16
/ PRIOR APPLICATION NUMBER: 60/263,479
/ PRIOR FILING DATE: 2001-01-23
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: FaSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 3010
/ TYPE: PRT
/ ORGANISM: Con 1 HCV isolate nucleic acid
US-10-467-000-1

```

```

Query Match          95.0%; Score 1887; DB 12; Length 3010;
Best Local Similarity 93.4%; Pred. No. 3,5e-182;

```

Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

QY 2 LITLSPYKVLARLIMWLYLITRVEAHLQWIPPLNVRGGRDAIILLTCVAHEBELIFD 61
DB 828 LITLSPHYKLFARLIMWLYFITRAEHLQWIPPLNVRGGRDAIILLTCHEBELIFT 887

QY 62 ITKLILATFGLMVAQAGITKVPYFRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 121
DB 888 ITKLILATFGLMVAQAGITKVPYFRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 947

QY 122 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARG 181
DB 948 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARG 1007

QY 182 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTATQ 241
DB 1008 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTATQ 1067

QY 242 SFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCTGS 301
DB 1068 SFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCTGS 1127

QY 302 SDLYVTRHADVIIVRRRGRSGSLSPRVSYLKSGSGGFLPCPSGHAAGIFRAAVCTR 361
DB 1128 SDLYVTRHADVIIVRRRGRSGSLSPRVSYLKSGSGGFLPCPSGHAAGIFRAAVCTR 1187

QY 362 GVAKAVDFIPVESMETTR 380
DB 1188 GVAKAVDFIPVESMETTR 1206

RESULT 10
US-09-919-901-4
Sequence 4, Application US/09919901
Publication No. US20030082518A1
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-02-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: :
US-09-919-901-4

Query Match 94.7%; Score 1881; DB 10; Length 1692;
Best Local Similarity 93.1%; Pred. No. 6,4e-182;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 182 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTATQ 241
DB 287 KEILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTATQ 346

QY 242 SFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCTGS 301
DB 347 SFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCTGS 406

QY 302 SDLYVTRHADVIIVRRRGRSGSLSPRVSYLKSGSGGFLPCPSGHAAGIFRAAVCTR 361
DB 407 SDLYVTRHADVIIVRRRGRSGSLSPRVSYLKSGSGGFLPCPSGHAAGIFRAAVCTR 466

QY 362 GVAKAVDFIPVESMETTR 380
DB 467 GVAKAVDFIPVESMETTR 485

RESULT 11
US-10-191-966-4
Sequence 4, Application US/10191966
Publication No. US20030175692A1
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT FILING DATE: 2002-07-10
PRIOR APPLICATION NUMBER: US/09/263,933
PRIOR FILING DATE: 1999-03-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: :
US-10-191-966-4

Query Match 94.7%; Score 1881; DB 14; Length 1692;
Best Local Similarity 93.1%; Pred. No. 6,4e-182;
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LITLSPYKVLARLIMWLYLITRVEAHLQWIPPLNVRGGRDAIILLTCVAHEBELIFD 61
DB 107 LITLSPYKVLARLIMWLYFITRAEHLQWIPPLNVRGGRDAIILLTCVAHEBELIFD 166

QY 62 ITKLILATFGLMVAQAGITKVPYFRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 121
DB 167 ITKLILATFGLMVAQAGITKVPYFRAQGLIRACMLVRKAGGHYQMAFMKLAALGT 226

QY 122 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARG 181
DB 227 YVYDHLTPLODMAHAGLRDLAAVEPVIFSDMEVKIITWGADTAACDIIISGLPVSARG 286

QY 182 REILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTATQ 241
DB 287 KEILGPDNFEQGGWRLAPITAYSOQTRGLGCIITSLGRDKNOVEGEVQVSTATQ 346

QY 242 SFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCTGS 301
DB 347 SFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDODLVGMQAPPGARSMTPTCTGS 406

QY 302 SDLYVTRHADVIIVRRRGRSGSLSPRVSYLKSGSGGFLPCPSGHAAGIFRAAVCTR 361
DB 407 SDLYVTRHADVIIVRRRGRSGSLSPRVSYLKSGSGGFLPCPSGHAAGIFRAAVCTR 466

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 47.9132 Seconds

(without alignments)
2075.771 Million cell updates/sec

Title: US-10-650-585-13

Perfect score: 1842
Sequence: 1 AHQWIPPLNRCGRDAIT.....RGVAKAVDFIPVSMETMR 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1842	100.0	352	ABG32186 HCV prote
2	1842	100.0	380	ABG32185 HCV prote
3	1842	100.0	393	ABG32184 HCV prote
4	1842	100.0	409	ABG32181 HCV prote
5	1778	96.5	341	ABG32187 HCV prote
6	1773	96.3	3010	AA82694 Partial H
7	1768	96.0	3010	AA86862 HCV prote
8	1767	95.9	2	AA86864 Hepatitis
9	1766	95.9	2201	ABG30591 Hepatitis
10	1766	95.9	2201	ABG30591 Hepatitis
11	1766	95.9	2201	ABG30591 Hepatitis
12	1766	95.9	2201	ABG30591 Hepatitis
13	1766	95.9	2201	ABG30591 Hepatitis
14	1766	95.9	2201	ABG30591 Hepatitis
15	1766	95.9	2201	ABG30591 Hepatitis
16	1766	95.9	2201	ABG30591 Hepatitis
17	1766	95.9	2201	ABG30591 Hepatitis
18	1766	95.9	2201	ABG30591 Hepatitis
19	1766	95.9	2201	ABG30591 Hepatitis
20	1766	95.9	2201	ABG30591 Hepatitis
21	1766	95.9	3010	ABG32458 Hepatitis
22	1766	95.9	3010	ABG32458 Hepatitis
23	1766	95.9	3010	ABG32451 Hepatitis
24	1766	95.9	3010	ABG32455 Hepatitis
25	1766	95.9	3010	ABG32457 Hepatitis

26	1766	95.9	3010	5	ABG32460	Abg32460 Hepatitis
27	1766	95.9	3010	5	ABG32461	Abg32461 Hepatitis
28	1766	95.9	3010	5	ABG32454	Abg32454 Hepatitis
29	1766	95.9	3011	5	ABG32456	Abg32456 Hepatitis
30	1763	95.7	2201	5	ABG30586	Abg30586 Hepatitis
31	1763	95.7	2201	5	ABG30589	Abg30589 Hepatitis
32	1763	95.7	2201	5	ABG30583	Abg30583 Hepatitis
33	1763	95.7	2201	5	ABG30588	Abg30588 Hepatitis
34	1762	95.7	2201	5	ABG30590	Abg30590 Hepatitis
35	1760	95.5	3010	5	ABG32452	Abg32452 Hepatitis
36	1759	95.5	2201	5	ABG30584	Abg30584 Hepatitis
37	1759	95.5	2201	5	ABG30602	Abg30602 Hepatitis
38	1759	95.5	3010	5	ABG32453	Abg32453 Hepatitis
39	1758	95.4	768	2	AA840223	AA840223 Recombin
40	1757	95.4	2307	3	AA870064	AA870064 Recombin
41	1757	95.3	2307	3	AA870065	AA870065 Recombin
42	1755	95.3	2307	3	AA870065	AA870065 Recombin
43	1748	94.9	3010	2	AA898022	AA898022 Infectiou
44	1748	94.9	3010	4	AA859174	AA859174 Protein e
45	1748	94.9	3010	4	AA831170	AA831170 Amino aci

ALIGNMENTS

RESULT 1
ID ABG32186 standard; protein; 352 AA.
XX
AC ABG32186;
XX
AC
XX
DT 05-NOV-2002 (first entry)
XX
DE HCV protease NS2/3 truncation mutant 855-1206.
XX
XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;
KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
KM hepatocellular carcinoma; anti-hepatitis C virus; lauryldithylamine oxide; LDHO;
KM chaotropic agent; mutant; mutein.
XX
XX Hepatitis C virus.
OS Synthetic.
OS
PN WO200248375-A2.
XX
XX 20-JUN-2002.
XX
PF 13-DEC-2001; 2001WO-CA001796.
XX
PR 15-DEC-2000; 2000US-0256031P.
XX
PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
PI Thiobaut D, Lamarre D, Maurice R, Pilote L, Pause A;
XX WPI; 2002-599511/64.
XX
PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.
XX
XX
XX Claim 41; Page 61-62; 67pp; English.
CC The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldithylamine oxide (LDHO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

DB 209 QTRGLGCIITSLTGRDRKQVEGEVQVSTATQSFLLATCVNGVCMVTFHGAGSKTLAEPK 268
 QY 241 GPITOMYTNVDODLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRGSRLLS 300
 DB 269 GPITOMYTNVDODLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRGSRLLS 328
 QY 301 PRPVSYLKGSSGGPILCPSGHVGIFPRAVCTRGVAKAVDFIPVSMETMR 352
 DB 329 PRPVSYLKGSSGGPILCPSGHVGIFPRAVCTRGVAKAVDFIPVSMETMR 380

RESULT 3
 ABG32184
 ID ABG32184 standard; protein; 393 AA.
 XX
 AC ABG32184;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE HCV protease NS2/3 truncation mutant 815-1206.
 XX
 KM HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KM chaotropic agent; mutant; mutein.
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 PN WO200248375-A2.
 XX
 PD 20-JUN-2002.
 XX
 PF 13-DEC-2001; 2001MO-CA001796.
 XX
 PR 15-DEC-2000; 2000US-0256031P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Thiabeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 XX
 DR WPI; 2002-599511/64.
 XX
 PT Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against Hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.
 XX
 PS Claim 41; Page 59-60; 67pp; English.
 XX

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (c) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32184; (3) producing (M1) a reformed, inactive HCV NS2/3 protease, involving isolating the isolated protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an

CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterization of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 815-1206
 CC (numbered relative to the full length NS2/3 protein)
 CC
 XX
 SQ Sequence 393 AA;
 XX
 Query Match 100.0%; Score 1842; DB 5; Length 393;
 Best Local Similarity 100.0%; Pred. No. 4.4e-10;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHLQVWIPPLNVRGGRDAIILLTCVHPELIDITKLLAIFGPMVLQAGITKYPYFVR 60
 DB 42 AHLQVWIPPLNVRGGRDAIILLTCVHPELIDITKLLAIFGPMVLQAGITKYPYFVR 101
 QY 61 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTCTYYDHTPLQDVAHAGLRPLAVAVEPV 120
 DB 102 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTCTYYDHTPLQDVAHAGLRPLAVAVEPV 161
 QY 121 IFSDEWVKITWGADTLAACGDIISGLPVARSRGREILLGPADNFEQGWRLAPITAYSQ 180
 DB 162 IFSDEWVKITWGADTLAACGDIISGLPVARSRGREILLGPADNFEQGWRLAPITAYSQ 221
 QY 181 QTRGLGCIITSLTGRDRKQVEGEVQVSTATQSFLLATCVNGVCMVTFHGAGSKTLAEPK 240
 DB 222 QTRGLGCIITSLTGRDRKQVEGEVQVSTATQSFLLATCVNGVCMVTFHGAGSKTLAEPK 281
 QY 241 GPITOMYTNVDODLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRGSRLLS 300
 DB 282 GPITOMYTNVDODLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPTRRRGSRLLS 341
 QY 301 PRPVSYLKGSSGGPILCPSGHVGIFPRAVCTRGVAKAVDFIPVSMETMR 352
 DB 342 PRPVSYLKGSSGGPILCPSGHVGIFPRAVCTRGVAKAVDFIPVSMETMR 393

RESULT 4
 ABG32181
 ID ABG32181 standard; protein; 409 AA.
 XX
 AC ABG32181;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE HCV protease NS2/3 (810-1206).
 XX
 KM HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KM chaotropic agent; mutant; mutein.
 OS Hepatitis C virus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT 398..409
 FT Peptide /note="Streptavidin tag"
 XX
 PN WO200248375-A2.
 XX
 PD 20-JUN-2002.
 XX
 PF 13-DEC-2001; 2001MO-CA001796.
 XX
 PR 15-DEC-2000; 2000US-0256031P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thiabeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 PI MPI: 2002-599511/64.
 XX N-PSDB; ABR90406.
 XX

PT Novel polypeptide for screening inhibitors of non-structural proteases
 useful as therapeutic agents against hepatitis C virus, comprises full
 length non-structural protease, or its truncation.
 PT

PS Claim 42; Fig 1B; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 residue amino acid 810 to 906, or having a minimal amino acid sequence
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 NS2/3 protease. Also included are (1) a composition (C) comprising an
 isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 its truncation or a mutated sequence, where the protease is in a solution
 comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 appearing as ABG32187; (3) producing (M1) a refolded, inactive HCV NS2/3
 protease, involving isolating the protease in the presence of a
 chaotropic agent, refolding the isolated protease by contacting it with a
 chaotropic agent, and LDAO in the presence of reduced concentration of the
 chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 protease, involving diluting refolded inactive NS2/3 protease in a medium
 containing an activation detergent to induce auto-cleavage of the NS2/3
 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 protease, involving incubating the active NS2/3 protease produced by M2
 for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 cleavage products or their fragments, and measuring the presence or
 absence of uncleaved NS2/3 protease, cleavage products or their fragments
 ; and (6) screening a potential inhibitor of auto-cleavage activity of an
 active NS2/3 protease, involving carrying out M3 in the presence of, or
 absence of the potential inhibitor, comparing the amount of uncleaved
 NS2/3 protease, cleavage products or their fragments. The protease is
 useful for detailed biochemical characterisation of the enzymes and in
 the development of in vitro assays for screening novel inhibitors of
 NS2/3 protease which are useful as therapeutic agents against HCV
 infection (which causes chronic liver disease, cirrhosis and end-stage
 liver disease. M1 is useful for high level production of protease. M2
 present sequence represents the NS2/3 (810-1206) protein, which has a C-
 terminal streptavidin tag

Sequence 409 AA;

Query Match 100.0%; Score 1842; DB 5; Length 409;
 Best Local Similarity 100.0%; Pred. No. 4.6e-170;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AHQVWIPPLNVRGRDAIILLTCVAPHEIFDITKLLAIFGLMVLQGITKVPFVR 60
 46 AHQVWIPPLNVRGRDAIILLTCVAPHEIFDITKLLAIFGLMVLQGITKVPFVR 105
 61 AGLIRACMLVRKAGHYVOMAFMKLAALTGTVVYHPLPLQWMAHGRDLAIVEPV 120
 106 AGLIRACMLVRKAGHYVOMAFMKLAALTGTVVYHPLPLQWMAHGRDLAIVEPV 165
 121 IFSDMEVKITMGADTAACGDIISGLPVSARGREILIGPADNFEQGMRLAIPITAYSQ 180
 166 IFSDMEVKITMGADTAACGDIISGLPVSARGREILIGPADNFEQGMRLAIPITAYSQ 225
 181 QTRGLIGCITSLTGRDNQVEGEVYSTATQSPFLATCNQGVWTFPHAGSSTTLAIPK 240
 226 QTRGLIGCITSLTGRDNQVEGEVYSTATQSPFLATCNQGVWTFPHAGSSTTLAIPK 285
 241 GPITOMYNNVODLVGMQAPPGARSMPTCTGSSDLVYTFHADVIFVRRGRSGSLLS 300
 286 GPITOMYNNVODLVGMQAPPGARSMPTCTGSSDLVYTFHADVIFVRRGRSGSLLS 345
 301 PRPVSYLKSSGGPLLCPSGHAAGVIFRAAVCTRGVAAVDFIPVESMETTMR 352

DB 346 PRPVSYLKSSGGPLLCPSGHAAGVIFRAAVCTRGVAAVDFIPVESMETTMR 397

RESULT 5
 ABG32187
 ID ABG32187 standard; protein; 341 AA.
 AC ABG32187;
 XX 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 866-1206.

XX HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;
 XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 XX hepatocellular carcinoma; anti-inflammatory; lauryldiethylamine oxide; LDAO;
 XX chaotropic agent; mutant; mutein.

OS Hepatitis C virus.
 OS Synthetic.
 XX WO200248375-A2.

PD 20-JUN-2002.

PF 13-DEC-2001; 2001WC-CA001796.

PR 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thiabeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 XX MPI: 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases
 useful as therapeutic agents against hepatitis C virus, comprises full
 length non-structural protease, or its truncation.
 PT

PS Claim 41; Page 62-63; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 residue amino acid 810 to 906, or having a minimal amino acid sequence
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 NS2/3 protease. Also included are (1) a composition (C) comprising an
 isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 its truncation or a mutated sequence, where the protease is in a solution
 comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 appearing as ABG32187; (3) producing (M1) a refolded, inactive HCV NS2/3
 protease, involving isolating the protease in the presence of a
 chaotropic agent, refolding the isolated protease by contacting it with a
 chaotropic agent, and LDAO in the presence of reduced concentration of the
 chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 protease, involving diluting refolded inactive NS2/3 protease in a medium
 containing an activation detergent to induce auto-cleavage of the NS2/3
 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 protease, involving incubating the active NS2/3 protease produced by M2
 for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 cleavage products or their fragments, and measuring the presence or
 absence of uncleaved NS2/3 protease, cleavage products or their fragments
 ; and (6) screening a potential inhibitor of auto-cleavage activity of an
 active NS2/3 protease, involving carrying out M3 in the presence of, or
 absence of the potential inhibitor, comparing the amount of uncleaved
 NS2/3 protease, cleavage products or their fragments. The protease is
 useful for detailed biochemical characterisation of the enzymes and in
 the development of in vitro assays for screening novel inhibitors of
 NS2/3 protease which are useful as therapeutic agents against HCV
 infection (which causes chronic liver disease, cirrhosis and end-stage
 liver disease. M1 is useful for high level production of protease. The

CC present sequence represents the NS2/3 truncation mutant 866-1206
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 341 AA;

Query Match 96.5%; Score 1778; DB 5; Length 341;

Best Local Similarity 100.0%; Pred. No. 5,9e-164; Mismatches 0; Indels 0; Gaps 0;

Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

12 VRGRSDAIIILTCVHPELFDITKLLAFGLPMTVAQITKYPYFVRAQGLIRACMLV 71

1 VRGRSDAIIILTCVHPELFDITKLLAFGLPMTVAQITKYPYFVRAQGLIRACMLV 60

72 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDLAVAVEPVIFSDMEVKIIT 131

61 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDLAVAVEPVIFSDMEVKIIT 120

132 WGAADTAACGDIISGLPVASARRGEIILGPADNFEQGMRLAPITAYSQOTRGLGCTIT 191

121 WGAADTAACGDIISGLPVASARRGEIILGPADNFEQGMRLAPITAYSQOTRGLGCTIT 180

192 SLTRDKQVGEVGVSTATQSFATCVGVCWTVFHGASGKTLGPKPITQMTNV 251

181 SLTRDKQVGEVGVSTATQSFATCVGVCWTVFHGASGKTLGPKPITQMTNV 240

252 QDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVTPRRRQDSRGSLSFRPVSYLKSS 311

241 QDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVTPRRRQDSRGSLSFRPVSYLKSS 300

312 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 352

301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 341

RESULT 6

AAR82694 ID AAR82694 standard; protein; 3010 AA.

AC AAR82694;

16-OCT-2003 (revised)

14-NOV-1996 (first entry)

Partial HCV non-structural polyprotein.

proteinase; hepatitis C virus; screening; inhibitor; proteolytic;

identification; cleavage.

Hepatitis C virus; Virus.

Key Location/Qualifiers

Protein 898..1233

FT /note="partial proteinase; see AAR82692"

FT Protein 992..1907

FT /note="partial proteinase; see AAR82693"

XX JP07184648-A.

XX 25-JUL-1995.

XX 05-FEB-1993; 93JP-00018854.

XX 07-FEB-1992; 92JP-00022657.

XX 18-SEP-1992; 92JP-00249240.

XX 04-DEC-1992; 92JP-00325303.

XX (KAEN/) KAENNO K.

XX (SUMO-) SUMITOMO METAL IND LTD.

XX (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.

XX WPI; 1995-287962/38.

XX N-PSDB; AAT03960.

PT An HCV proteinase active substance - which has activity as an anti-HCV
 PT agent and can be used to screen for proteinase inhibitors.

XX Disclosure; Page 39-48; 52pp; Japanese.

CC The present sequence is a partial Hepatitis C Virus (HCV) polyprotein
 CC from the non-structural region. Partial proteinase sequences (AAR82692-
 CC 93) are contained within this sequence. The proteinases can be used as
 CC anti-HCV agents. They can also be used to screen cpts. for their ability
 CC to inhibit their proteolytic activity. In this way proteinase inhibitors
 CC can be identified. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 3010 AA;

Query Match 96.3%; Score 1773; DB 2; Length 3010;

Best Local Similarity 94.6%; Pred. No. 4e-162; Mismatches 7; Indels 0; Gaps 0;

Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

1 AHQVWIPPLNRRGRDAIILTCVHPELFDITKLLAFGLPMTVAQITKYPYFVR 60

855 AHQVWIPPLNRRGRDAIILTCVHPELFDITKLLAFGLPMTVAQITKYPYFVR 914

61 AAGLRACMLVRKAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDLAVAVEPV 120

915 AAGLRACMLVRKAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDLAVAVEPV 974

121 IFSDEVKIITWGAADTAACGDIISGLPVASARRGEIILGPADNFEQGMRLAPITAYSQ 180

975 IFSDEVKIITWGAADTAACGDIISGLPVASARRGEIILGPADNFEQGMRLAPITAYSQ 1034

181 QTRGLGCTITSLTRDKQVGEVGVSTATQSFATCVGVCWTVFHGASGKTLGPK 240

1035 QTRGLGCTITSLTRDKQVGEVGVSTATQSFATCVGVCWTVFHGASGKTLGPK 1094

241 GPITQMTNVDPDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVTPRRRQDSRGSLS 300

1095 GPITQMTNVDPDLVGMQAPPGARSMPTCTGSSDLVYVTRHADVTPRRRQDSRGSLS 1154

301 PRPVSYLKSSGSGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 352

1155 PRPVSYLKSSGSGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 7

AAR86622 ID AAR86622 standard; protein; 3010 AA.

AC AAR86622;

16-OCT-2003 (revised)

16-OCT-1995 (first entry)

HCV protein cleavable with new serine proteinase.

proteinase; serine; cleavage; hepatitis C virus; HCV.

Hepatitis C virus; Virus.

Key Location/Qualifiers

FT Cleavage-site 2419..2420

FT /note="Serine protease cleavage site"

XX JP06315377-A.

XX 15-NOV-1994.

XX 06-MAY-1993; 93JP-00105666.

XX 06-MAY-1993; 93JP-00105666.

XX (KAEN/) KAENNO K.

XX (SUMO-) SUMITOMO METAL IND LTD.

XX (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.

XX WPI; 1995-032330/05.
 DR N-PSDB; AA080498.
 XX
 PT New HCV-originated proteinase active substance - used for site-specific
 cleavage by an intermolecular reaction and the purification thereof.
 XX
 PS Disclosure; Page 10-13; 23pp; Japanese.
 CC
 CC This protein from HCV (hepatitis C virus) (encoded by AA080498) is
 cleaved between amino acids 2419 and 2420, by a new serine protease,
 CC contg. the sequence of AAR68621. The proteinase is purified as a fused
 CC product with the dihydrofolate reductase protein by using a methotrexate
 CC column. It can be used for the development of an inhibitor for HCV
 CC proteinase. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 3010 AA;
 Query Match 96.0%; Score 1769; DB 2; Length 3010;
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;
 Matches 332; Conservative 12; Mismatches 8; Indels 0; Gaps 0;
 QY 1 AHIQWIPPLNVNNGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITRVYFVR 60
 DB 855 AHIQWIPPLNVNNGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITRVYFVR 914
 QY 61 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAVAVEPV 120
 DB 915 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAVAVEPV 974
 QY 121 IFSMEVKIITWGAADTAACDIIISGLPVSARREIILGPADNFBEGQWRLAIPITAYSQ 180
 DB 975 VFSMETKLIITWGAADTAACDIIISGLPVSARREIILGPADNFBEGQWRLAIPITAYSQ 1034
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNGVCMVTFGAGSKTLAAGPK 240
 DB 1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNGVCMVTFGAGSKTLAAGPK 1094
 QY 241 GPITQMTNVDQDLYGMAPEPARSMTPTCTGSSSDIYLVTRHADVIPIRRRGDSRGLIS 300
 DB 1095 GPITQMTNVDQDLYGMAPEPARSMTPTCTGSSSDIYLVTRHADVIPIRRRGDSRGLIS 1154
 QY 301 PRPISYIKGSSGGPLCPSGHVGIFPAVCTRGVAKADFIIVESMETTR 352
 DB 1155 PRPISYIKGSSGGPLCPSGHVGIFPAVCTRGVAKADFIIVESMETTR 1206

RESULT 8
 AAR6864 standard; protein: 3010 AA.
 ID AAR6864
 XX
 AC AAR6864;
 XX
 DT 06-DEC-1995 (first entry)
 XX
 DE Hepatitis C virus RNA helicase.
 XX
 XX Hepatitis C virus, HCV; non-A non-B; helicase gene; RNA helicase;
 KM baculovirus; recombinant production.
 XX
 OS Hepatitis C virus.
 XX
 XX
 FT Key
 FT Location/Qualifiers
 FT 196..198
 FT /label=N-linked glycosylation site
 FT 209..211
 FT /label=N-linked glycosylation site
 FT 234..236
 FT /label=N-linked glycosylation site
 FT 250..252
 FT /label=N-linked glycosylation site
 FT 305..307
 FT /label=N-linked glycosylation site

FT Region 325..327
 FT /label=N-linked glycosylation site
 FT Region 417..419
 FT /label=N-linked glycosylation site
 FT Region 423..425
 FT /label=N-linked glycosylation site
 FT Region 430..432
 FT /label=N-linked glycosylation site
 FT Region 448..450
 FT /label=N-linked glycosylation site
 FT Region 532..534
 FT /label=N-linked glycosylation site
 FT Region 556..558
 FT /label=N-linked glycosylation site
 FT Region 576..578
 FT /label=N-linked glycosylation site
 FT Region 623..625
 FT /label=N-linked glycosylation site
 FT Region 645..647
 FT /label=N-linked glycosylation site
 FT Region 1213..1215
 FT /label=N-linked glycosylation site
 FT Region 1255..1257
 FT /label=N-linked glycosylation site
 FT Region 2041..2043
 FT /label=N-linked glycosylation site
 FT Region 2077..2079
 FT /label=N-linked glycosylation site
 FT Region 2240..2242
 FT /label=N-linked glycosylation site
 FT Region 2788..2790
 FT /label=N-linked glycosylation site

PN JP06319583-A.
 XX
 PD 22-NOV-1994.
 XX
 PR 18-SEP-1992; 92JP-00249241.
 XX
 PR 18-SEP-1992; 92JP-00249241.
 XX
 PA (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.
 XX
 DR WPI; 1995-040330/06.
 XX
 DR N-PSDB; AA081559.
 XX
 XX
 XX Of hepatitis C virus helicase gene in baculovirus - useful for large
 PT scale prodn. of RNA helicase.
 XX
 XX
 PS Claim 1; Fig 1-4; 9pp; Japanese.
 CC
 CC AA081559 encodes AAR6864 hepatitis C virus (HCV) RNA helicase. The DNA
 CC was used in the construction of an expression vector, which was used to
 CC transform a baculovirus host. The transformed baculovirus could then be
 CC used for the recombinant prodn. of HCV RNA helicase
 XX
 SQ Sequence 3010 AA;
 Query Match 95.9%; Score 1767; DB 2; Length 3010;
 Best Local Similarity 94.3%; Pred. No. 1.5e-161;
 Matches 332; Conservative 12; Mismatches 8; Indels 0; Gaps 0;
 QY 1 AHIQWIPPLNVNNGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITRVYFVR 60
 DB 855 AHIQWIPPLNVNNGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITRVYFVR 914
 QY 61 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAVAVEPV 120
 DB 915 AOGILRACMLVRKAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGLDLAVAVEPV 974
 QY 121 IFSMEVKIITWGAADTAACDIIISGLPVSARREIILGPADNFBEGQWRLAIPITAYSQ 180
 DB 975 VFSMETKLIITWGAADTAACDIIISGLPVSARREIILGPADNFBEGQWRLAIPITAYSQ 1034

CC the claims of the invention
 XX Sequence 2201 AA;
 Query Match 95.9%; Score 1766; DB 5; Length 2201;
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAEPK 240
 DB 1035 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAEPK 1094

QY 241 GPTTQMTNTVDODLVGMQAPRGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRSLLS 300
 DB 1095 APTTQMTNTVDODLVGMQAPRGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRSLLS 1154

QY 301 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 352
 DB 1155 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

RESULT 9
 ABG30601
 ID ABG30601 standard; protein; 2201 AA.
 AC ABG30601;
 XX
 DT 21-OCT-2002 (first entry)
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
 XX
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 882
 FT Misc-difference 2183 /label= Arg, Lys
 FT /note= "Wild type Met substituted by Thr"
 FT
 XX
 FT
 XX
 PN WO200252015-A2.
 PD 04-JUL-2002.
 XX
 PD 20-DEC-2001; 2001MO-CA001843.
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukolj G, Pause A;
 XX
 DR WPI; 2002-575382/61.
 XX
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 PS Claim 3; Page; 140pp; English.
 XX
 CC The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
 CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G12042/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis
 CC C virus replicon AAGK12 and contains the viral protease NS2/3, protease
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
 CC This sequence does not appear in the specification but has been created
 CC from the wild type sequence shown in ABG30580 using information given in

CC the claims of the invention
 XX Sequence 2201 AA;
 Query Match 95.9%; Score 1766; DB 5; Length 2201;
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHQWVPIPLNVRGGDAIILITCAVHPELLFDITKLLAIFGPMVLAQGITKVPYVR 60
 DB 46 AHQWVPIPLNVRGGDAIILITCAVHPELLFDITKLLAIFGPMVLAQGITKVPYVR 105

QY 61 AAGLIRACMLVAKAGGHVQVAFMKLAALITGYDDHTPLQDVAAHAGRLAVAVEPV 120
 DB 106 AAGLIRACMLVAKAGGHVQVAFMKLAALITGYDDHTPLQDVAAHAGRLAVAVEPV 165

QY 121 IFSDMEVKIITWGAFTAAAGDIIISGLPVASARGREIILGPADNFEQGGRLIAPITAYSQ 180
 DB 166 VPSDMETKVIITWGAFTAAAGDIIISGLPVASARGREIILGPADNFEQGGRLIAPITAYSQ 225

QY 181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAEPK 240
 DB 226 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCNGVCWTFVHGAGSKTLAEPK 285

QY 241 GPTTQMTNTVDODLVGMQAPRGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRSLLS 300
 DB 286 GPTTQMTNTVDODLVGMQAPRGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRSLLS 345

QY 301 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 352
 DB 346 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 397

RESULT 10
 ABG30591
 ID ABG30591 standard; protein; 2201 AA.
 AC ABG30591;
 XX
 DT 21-OCT-2002 (first entry)
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.
 XX
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 751
 FT Misc-difference /note= "Wild type Ser substituted by Gly"
 FT Misc-difference 882 /label= Arg, Lys
 FT
 XX
 FT
 XX
 PN WO200252015-A2.
 PD 04-JUL-2002.
 XX
 PD 20-DEC-2001; 2001MO-CA001843.
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukolj G, Pause A;
 XX
 DR WPI; 2002-575382/61.
 XX
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.

PS Claim 3; Page; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)

CC polynucleotide molecule comprising a 5'-non translated region (NTR),

CC where guanine at position 1 is substituted for adenine, a HCV polypeptide

CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-

CC replicating hepatitis C virus (HCV) RNA molecule is useful for evaluating

CC potential inhibitors of HCV replication. The HCV RNA molecule is also

CC useful for efficiently establishing cell culture replication. The self-

CC replicating polynucleotide molecule contains a 5'-NTR, where G at

CC position 1 is substituted for A, and therefore provides an alternative to

CC existing systems comprising a self-replicating HCV RNA molecule that, in

CC conjunction with mutations in the HCV non-structural region, such as the

CC G2042C/R mutations, transduces and/or replicates with greater

CC efficiency. This amino acid sequence represents a mutant of the hepatitis

CC C virus replicon ApgK12 and contains the viral protease NS2/3, protease

CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:

CC This sequence does not appear in the specification but has been created

CC from the wild type sequence shown in ABG30580 using information given in

CC the claims of the invention

XX

XX Sequence 2201 AA;

XX

XX Query Match 95.3%; Score 1766; DB 5; Length 2201;

XX Best Local Similarity 94.3%; Pred. No. 1.2e-161;

XX Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHHQWIPPLNVRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 60

DB 46 AHHQWIPPLNVRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 105

QY 61 AAGLIRACMLVRKAAAGHYVQMAFMKLAALGTGVVYDHLPIODMAHAGRLDAVAVEPV 120

DB 106 AAGLIRACMLVRKAAAGHYVQMAFMKLAALGTGVVYDHLPIODMAHAGRLDAVAVEPV 165

QY 121 IFSDMEVKIITMGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180

DB 166 VFSDEMETKVIITMGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 225

QY 181 QTRGLGCIITSLTGRBNQVEGEVQVSTATOSFLATCNQVCMVTHGASGKTLAAGPK 240

DB 226 QTRGLGCIITSLTGRBNQVEGEVQVSTATOSFLATCNQVCMVTHGASGKTLAAGPK 285

QY 241 GPITQMTNNVQDVLGMAOPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 300

DB 286 GPITQMTNNVQDVLGMAOPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 345

QY 301 PRPVSYLKSSGGPILCPSGHAAGIIPRAAVCTRGVAKAVDFIVESMETTMR 352

DB 346 PRPVSYLKSSGGPILCPSGHAAGIIPRAAVCTRGVAKAVDFIVESMETTMR 397

DB

RESULT 11

ABG30600

ID ABG30600 standard; protein; 2201 AA.

XX

XX ABG30600;

XX

XX 21-OCT-2002 (first entry)

XX

XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

DE

XX

XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

XX

XX cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutcin.

XX

XX Hepatitis C virus.

OS

XX Synthetic.

XX

XX Key Location/Qualifiers

XX

XX Msc-difference 882

XX

XX Msc-difference 1357

XX

XX Msc-difference 1357

XX

XX /label= Arg, Lys

XX

XX /note= "Wild type Pro substituted by Leu"

XX

XX W0200252015-A2.

XX

XX 04-JUL-2002.

XX

XX 20-DEC-2001; 2001WO-CAN001843.

XX

XX 22-DEC-2000; 2000US-0257857P.

XX

XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX

XX Kukulj G, Pause A;

XX

XX MPI, 2002-575382/61.

XX

XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which

XX

XX possess enhanced transduction or replication efficiency, useful for

XX

XX evaluating potential inhibitors of HCV replication.

XX

XX Claim 3; Page; 140pp; English.

XX

XX The invention describes a self-replicating hepatitis C virus (HCV)

XX

XX polynucleotide molecule comprising a 5'-non translated region (NTR),

XX

XX where guanine at position 1 is substituted for adenine, a HCV polypeptide

XX

XX region coding for a HCV polypeptide; and a 3'-NTR region. The self-

XX

XX replicating hepatitis C virus (HCV) RNA molecule is useful for evaluating

XX

XX potential inhibitors of HCV replication. The HCV RNA molecule is also

XX

XX useful for efficiently establishing cell culture replication. The self-

XX

XX replicating polynucleotide molecule contains a 5'-NTR, where G at

XX

XX position 1 is substituted for A, and therefore provides an alternative to

XX

XX existing systems comprising a self-replicating HCV RNA molecule that, in

XX

XX conjunction with mutations in the HCV non-structural region, such as the

XX

XX G(2042)C/R mutations, transduces and/or replicates with greater

XX

XX efficiency. This amino acid sequence represents a mutant of the hepatitis

XX

XX C virus replicon ApgK12 and contains the viral protease NS2/3, protease

XX

XX complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:

XX

XX This sequence does not appear in the specification but has been created

XX

XX from the wild type sequence shown in ABG30580 using information given in

XX

XX the claims of the invention

XX

XX

XX Sequence 2201 AA;

XX

XX

XX Query Match 95.3%; Score 1766; DB 5; Length 2201;

XX

XX Best Local Similarity 94.3%; Pred. No. 1.2e-161;

XX

XX Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHHQWIPPLNVRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 60

DB 46 AHHQWIPPLNVRGGRDAIILTCVAHPELIFDITKLLAIFGPMVLOAGITKVPYFVR 105

QY 61 AAGLIRACMLVRKAAAGHYVQMAFMKLAALGTGVVYDHLPIODMAHAGRLDAVAVEPV 120

DB 106 AAGLIRACMLVRKAAAGHYVQMAFMKLAALGTGVVYDHLPIODMAHAGRLDAVAVEPV 165

QY 121 IFSDMEVKIITMGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180

DB 166 VFSDEMETKVIITMGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 225

QY 181 QTRGLGCIITSLTGRBNQVEGEVQVSTATOSFLATCNQVCMVTHGASGKTLAAGPK 240

DB 226 QTRGLGCIITSLTGRBNQVEGEVQVSTATOSFLATCNQVCMVTHGASGKTLAAGPK 285

QY 241 GPITQMTNNVQDVLGMAOPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 300

DB 286 GPITQMTNNVQDVLGMAOPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 345

QY 301 PRPVSYLKSSGGPILCPSGHAAGIIPRAAVCTRGVAKAVDFIVESMETTMR 352

DB 346 PRPVSYLKSSGGPILCPSGHAAGIIPRAAVCTRGVAKAVDFIVESMETTMR 397

DB

RESULT 12

ABG30581

ID ABG30581 standard; protein; 2201 AA.
 AC ABG30581;
 XX 21-OCT-2002 (first entry)
 XX
 XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.
 DE
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B.
 XX Hepatitis C virus.
 OS
 XX WO200252015-A2.
 XX
 XX 04-JUL-2002.
 XX
 XX 20-DEC-2001; 2001WO-CA001843.
 XX
 XX 22-DEC-2000; 2000US-0257857P.
 PR
 XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.
 PA
 XX Kukulj G, Pause A;
 XX
 XX WPI; 2002-575382/61.
 DR N-PSDB; ABK88573.
 DR
 XX
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 PS Disclosure; Page 49-58; 140pp; English.
 XX
 CC The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
 CC replicon APK12 and contains the viral protease NS2/3, protease complex
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
 CC
 XX
 SQ Sequence 2201 AA;
 XX
 Query Match 95.9%; Score 1766; DB 5; Length 2201;
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
 QY 1 AHHQVMTPLNVRGGRDAIILTCVAHPELIFDTIKLLAIFGLMVLQGITKVPYFVR 60
 DB 46 AHHQVMTPLNVRGGRDAIILTCVAHPELIFDTIKLLAIFGLMVLQGITKVPYFVR 105
 QY 61 AAGIIRACMVRKAGGVOMAFMKLAALTGTVVYHPLPDMWAGRDIAVANEV 120
 DB 106 AAGIIRACMVRKAGGVOMAFMKLAALTGTVVYHPLPDMWAGRDIAVANEV 165
 QY 121 IFSDEMEKIIITWGAADTAACGDIISGLPVSARSGREIILGPADNEEGGWELLAPITAYSQ 180
 DB 166 VFSDEMEKIIITWGAADTAACGDIISGLPVSARSGREIILGPADNEEGGWELLAPITAYSQ 225
 QY 181 QTRGLGCIITSLTGRGKNOVEGVVSTATOSFLATCNGVQWTFPHAGSKTLAAGPK 240
 DB 226 QTRGLGCIITSLTGRGKNOVEGVVSTATOSFLATCNGVQWTFPHAGSKTLAAGPK 285
 QY 241 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVLPVRRRDSRGSILS 300

DB 286 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVLPVRRRDSRGSILS 345
 QY 301 PREVSYLKSGSGGPLLCPGSAVGIFFRAVCTRGVAKAVDFPVSMEETMR 352
 DB 346 PREVSYLKSGSGGPLLCPGSAVGIFFRAVCTRGVAKAVDFPVSMEETMR 397
 RESULT 13
 ID ABG30593
 AC ABG30593 standard; protein; 2201 AA.
 XX
 XX ABG30593;
 XX
 XX 21-OCT-2002 (first entry)
 XX
 XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #4.
 DE
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
 XX Hepatitis C virus.
 OS
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 882
 FT /label= Arg, Lys
 FT Misc-difference 892 /note= "Wild type Leu substituted by Phe"
 FT
 XX
 XX WO200252015-A2.
 XX
 XX 04-JUL-2002.
 XX
 XX 20-DEC-2001; 2001WO-CA001843.
 XX
 XX 22-DEC-2000; 2000US-0257857P.
 XX
 XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.
 PA
 XX Kukulj G, Pause A;
 XX
 XX WPI; 2002-575382/61.
 DR
 XX
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 PS Claim 3; Page; 140pp; English.
 XX
 CC The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis
 CC C virus replicon APK12 and contains the viral protease NS2/3, protease
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
 CC This sequence does not appear in the specification but has been created
 CC from the wild type sequence shown in ABG30580 using information given in
 CC the claims of the invention
 CC
 XX
 SQ Sequence 2201 AA;
 XX
 Query Match 95.9%; Score 1766; DB 5; Length 2201;
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;

Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60
 DB 46 AHLQWIPPLNVNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 105
 QY 61 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVTHLTPLODMAHAGRLDAVAPEV 120
 DB 106 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVTHLTPLODMAHAGRLDAVAPEV 165
 QY 121 IFSDEMKVITWGAADTAACGDIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 180
 DB 166 VFSDEMKVITWGAADTAACGDIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 225
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNVCWTVFPGAGSKTLAEPK 240
 DB 226 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNVCWTVFPGAGSKTLAEPK 285
 QY 241 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIIVRRRGSRSLS 300
 DB 286 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIIVRRRGSRSLS 345
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETMR 352
 DB 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETMR 397

RESULT 14
 ABG30582
 ID ABG30582 standard; protein; 2201 AA.
 AC ABG30582;
 XX
 DT 21-OCT-2002 (first entry)
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.
 XX
 KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 XX cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; muten.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 882
 FT /note= "Wild type Lys substituted by Lys or Arg"
 FT Misc-difference 1233
 FT /note= "Wild type Gly substituted by Cys"
 XX
 PN W0200252015-A2.
 XX
 PD 04-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-CA001843.
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukulj G, Pause A;
 XX
 PI WPI: 2002-575382/61.
 DR N-PSDB; ABR8574.
 XX
 XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 PS Disclosure; Page 59-69; 140pp; English.
 XX
 CC The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-

CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
 CC replicon Apgk12 and contains the viral protease NS2/3, protease complex
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this
 CC sequence has been created from replicon Apgk12 shown in ABG30581

Sequence 2201 AA;

Query Match 95.9%; Score 1766; DB 5; Length 2201;
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60
 DB 46 AHLQWIPPLNVNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 105
 QY 61 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVTHLTPLODMAHAGRLDAVAPEV 120
 DB 106 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVTHLTPLODMAHAGRLDAVAPEV 165
 QY 121 IFSDEMKVITWGAADTAACGDIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 180
 DB 166 VFSDEMKVITWGAADTAACGDIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 225
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNVCWTVFPGAGSKTLAEPK 240
 DB 226 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNVCWTVFPGAGSKTLAEPK 285
 QY 241 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIIVRRRGSRSLS 300
 DB 286 GPITQMTNNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIIVRRRGSRSLS 345
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETMR 352
 DB 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETMR 397

RESULT 15
 ABG30580
 ID ABG30580 standard; protein; 2201 AA.
 AC ABG30580;
 XX
 DT 21-OCT-2002 (first entry)
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.
 XX
 KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 XX cell culture replication; NS2/3; NS3/4; NS3; NS5B.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 882
 FT /note= "Encoded by ARG"
 FT
 PN W0200252015-A2.
 XX
 PD 04-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-CA001843;
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX

PI Kukolj G, Pause A;

XX WPI; 2002-575382/61.

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.

XX Disclosure; Page 69-74; 140pp; English.

CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)/C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC replicon Apgk12 and contains the viral protease NS2/3, protease complex
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B

XX Sequence 2201 AA;

SQ Query Match 95.9%; Score 1766; DB 5; Length 2201;

Best Local Similarity 94.3%; Pred. No. 1.2e-161;

Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHIQWMIPLNVRGGRDAIILITCAVPEELIPDITKLLAIFGPIWLOAGITKVPYFVR 60
DB 46 AHIQWMIPLNVRGGRDAIILITCAVPEELIPDITKLLAIFGPIWLOAGITKVPYFVR 105
QY 61 AOGIIRACMLVRKAGGHVYQMAFMKLAALTGVYVDHLPLODMWAGLRDLAVAVEPV 120
DB 106 AHGIRACMLVRKAGGHVYQMAFMKLAALTGVYVDHLPLODMWAGLRDLAVAVEPV 165
QY 121 IFSDMEYKITTGADTAAAGDIISGLPVSARRGEIILGPADNPEGQWRLLAPITAYSQ 180
DB 166 VFSDMEYKITTGADTAAAGDIISGLPVSARRGEIILGPADNPEGQWRLLAPITAYSQ 225
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSPATOSFLATCVNGVCMTFHGAGSKTLAPK 240
DB 226 QTRGLGCIITSLTGRDNQVEGEVQVSPATOSFLATCVNGVCMTFHGAGSKTLAPK 285
QY 241 GPITQMTNVDQILVGMQAPPGASMTPTCGSSDLYLVRHADVIPVRRRGRSGSLIS 300
DB 286 GPITQMTNVDQILVGMQAPPGASMTPTCGSSDLYLVRHADVIPVRRRGRSGSLIS 345
QY 301 PRPVSYLKSSGGGELCPGSHAVGIFRAAVCTRGVAKAVDFIVESMETMR 352
DB 346 PRPVSYLKSSGGGELCPGSHAVGIFRAAVCTRGVAKAVDFIVESMETMR 397

Search completed: May 6, 2004, 09:30:47
Job time : 48.9132 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 6, 2004, 09:22:36 ; Search time 11.4639 Seconds
(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-13

Sequence: 1 AHQVWIPPLNVRGRDAII.....RGVAKAVDFIPVSMETTVR 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR:78:*
2: PIR:1:*
3: PIR:2:*
4: PIR:3:*
5: PIR:4:*
6: PIR:5:*
7: PIR:6:*
8: PIR:7:*
9: PIR:8:*
10: PIR:9:*
11: PIR:10:*
12: PIR:11:*
13: PIR:12:*
14: PIR:13:*
15: PIR:14:*
16: PIR:15:*
17: PIR:16:*
18: PIR:17:*
19: PIR:18:*
20: PIR:19:*
21: PIR:20:*
22: PIR:21:*
23: PIR:22:*
24: PIR:23:*
25: PIR:24:*
26: PIR:25:*
27: PIR:26:*
28: PIR:27:*
29: PIR:28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1773	96.3	3010	1 GNMVUJ	genome polyprotein
2	1766	95.9	3010	1 A45573	genome polyprotein
3	1749	95.0	3010	1 GNMVUW	genome polyprotein
4	1717	93.2	3010	1 S18030	genome polyprotein
5	1699	92.2	3010	1 GNMVUC	genome polyprotein
6	1616	87.7	3011	1 GNMVWC	genome polyprotein
7	1610	87.4	3011	1 S40770	genome polyprotein
8	1605	87.1	3011	1 GNMVCH	genome polyprotein
9	1401	76.1	3014	1 JCS570	genome polyprotein
10	1303	70.7	3033	1 GNMVU8	genome polyprotein
11	1301	70.6	3033	1 JQ1303	genome polyprotein
12	1301	72.2	3005	1 T08841	polyprotein - dour
13	1342.5	18.6	2970	1 T08839	polyprotein - marm
14	102.5	5.6	660	1 VHMWH2	structural protein
15	102.5	5.6	692	1 H71426	hypothetical prote
16	101	5.5	564	1 S36637	signal recognition
17	101	5.5	600	1 E46642	DNA-directed DNA p
18	100.5	5.5	353	1 G87392	conserved hypochet
19	99	5.4	399	1 AH3038	conserved hypochet
20	99	5.4	399	1 C98247	hypothetical 50.8k
21	97.5	5.3	1085	1 T03521	cobn protein homol
22	95.5	5.2	470	1 TC4098	tetracycline 6-hyd
23	94.5	5.1	2796	1 UC4743	fatty-acid synthas
24	93.5	5.1	1380	1 T18309	receptor-adenylate
25	93	5.0	7463	1 T36248	CDA peptide synthet
26	92.5	5.0	706	1 S33761	transferrin precu
27	92.5	5.0	716	1 G93612	hypothetical prote
28	92	5.0	659	1 B44212	structural protein
29	91.5	5.0	3414	1 GNMVVE	genome polyprotein

30	91	4.9	504	2 A84212	hypothetical prote
31	90.5	4.9	868	2 H81775	aconitate hydratase
32	90.5	4.9	871	2 H72597	hypothetical prote
33	90.5	4.9	3069	2 H70656	fatty-acid synthase
34	89.5	4.9	961	2 AE0375	probable cation-tr
35	89	4.8	446	2 AP1509	conserved hypochet
36	88.5	4.8	652	2 C70688	probable nitrate r
37	88.5	4.8	3076	2 A87058	fatty acid synthase
38	87.5	4.8	3412	1 GNMVUB	genome polyprotein
39	86.5	4.7	347	2 S44167	malate dehydrogena
40	86.5	4.7	451	2 H82044	alpha-dicarbonylate
41	86.5	4.7	470	1 NM1VW8	exo-alpha-sialidas
42	86.5	4.7	707	2 D84154	cadmium-transport
43	86	4.7	338	1 S34984	UDPglucose 4-epime
44	85.5	4.6	348	1 H70549	probable pdhb prot
45	85	4.6	470	1 NM1V9	exo-alpha-sialidas

ALIGNMENTS

RESULT 1
GNMVUJ
genome polyprotein - hepatitis C virus (strain J)
N:contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jun-2001
R/Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A>Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A:Reference number: A39253; MUID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:ID90208; NID:9221610; PIDN:BAA14233.1; PID:9221611
R/Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A>Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A:Reference number: PS0085
A:Accession: PS0085
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KAT>
A:Experimental source: Japanese isolate
A:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-389/Product: major envelope protein E #status predicted <MEB>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F:1230-1237/Product: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1662/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,27

Query Match	96.3%	Score 1773;	DB 1;	Length 3010;
Best Local Similarity	94.6%	Pred. No. 6.8e-14;		
Matches 333;	Conservative 12;	Mismatches 7;	Indels 0;	Gaps 0;
DB	1	855	61	915
AGGLIRACGLVAKVKAAGGVQVAFMKLAALGTGVYDHLTPLOQVNAHGLRDLAAVAEPV 120				
AGGLIRACGLVAKVKAAGGVQVAFMKLAALGTGVYDHLTPLOQVNAHGLRDLAAVAEPV 974				

QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 180
 Db 975 VFSDMETKILITWGAADTAACDIIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 1034
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVNSTATOSPLATCNGVCWTFHAGSGKTLAGPK 240
 Db 1035 QTRGLGCIITSLTGRKNOVEGEVQVNSTATOSPLATCNGVCWTFHAGSGKTLAGPK 1094
 QY 241 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGDSRGLLS 300
 Db 1095 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGDSRGLLS 1154
 QY 301 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 352
 Db 1155 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 2
 A45573
 genome polyprotein - hepatitis C virus (strain JT)
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C/Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
 Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: B
 A:Reference number: A45573; MUID:92295714; PMID:1318627
 A:Accession: A45573
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-3010 <TRAN>
 A/Cross-references: GB:D11168; GB:D01171; NID:Q221612; PIDD:BA01943.1; PIDD:Q221613
 A:Experimental source: HCV-JT
 A:Note: Sequence extracted from NCBI bacbone (NCBIN:106206, NCBIP:106207)
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:116-191/Product: major envelope protein E #status predicted <ME>
 F:192-389/Product: nonstructural protein NS1 #status predicted <NS1>
 F:390-729/Product: nonstructural protein NS2 #status predicted <NS2>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis virus #status predicted <NS3>
 F:1230-1237/Product: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 95.9%; Score 1766; DB 1; Length 3010;
 Best Local Similarity 94.9%; Pred. No. 2,7e-140;
 Matches 334; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMYLQAGITVPYFVR 60
 Db 855 AHLQWIPPLNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMYLQAGITVPYFVR 914
 QY 61 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTIVYDHLTPLOMAHAGLRLDAVAEPV 120
 Db 915 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTIVYDHLTPLOMAHAGLRLDAVAEPV 974
 QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 180
 Db 975 VFSDMETKILITWGAADTAACDIIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 1034
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVNSTATOSPLATCNGVCWTFHAGSGKTLAGPK 240
 Db 1035 QTRGLGCIITSLTGRKNOVEGEVQVNSTATOSPLATCNGVCWTFHAGSGKTLAGPK 1094
 QY 241 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGDSRGLLS 300

Db 1095 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGDSRGLLS 1154
 QY 301 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 352
 Db 1155 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 3
 GNNVTV
 genome polyprotein - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C/Accession: A40244
 R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
 Virology 188, 102-113, 1992
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449
 A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 A/Cross-references: GB:M64754
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:11-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <ME>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis virus #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 95.0%; Score 1749; DB 1; Length 3010;
 Best Local Similarity 92.9%; Pred. No. 7.2e-139;
 Matches 327; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMYLQAGITVPYFVR 60
 Db 855 AHLQWIPPLNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLMYLQAGITVPYFVR 914
 QY 61 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTIVYDHLTPLOMAHAGLRLDAVAEPV 120
 Db 915 AAGLIRACMLVRKKAAGHYVQMAFMKLAALTGTIVYDHLTPLOMAHAGLRLDAVAEPV 974
 QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 180
 Db 975 VFSDMETKILITWGAADTAACDIIISGLPVSARSGREIILGPADNFEQGWLLAPITAYSQ 1034
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVNSTATOSPLATCNGVCWTFHAGSGKTLAGPK 240
 Db 1035 QTRGLGCIITSLTGRKNOVEGEVQVNSTATOSPLATCNGVCWTFHAGSGKTLAGPK 1094
 QY 241 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGDSRGLLS 300
 Db 1095 GPITOMYTNVDODLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRRGDSRGLLS 1154
 QY 301 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 352
 Db 1155 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 4
 S18030
 genome polyprotein - hepatitis C virus (isolate JKI)

N/Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
A/Variety: isolate UK1
C/Date: 19-May-2000 #sequence revision 19-May-2000 #text_change 23-Mar-2001
C/Accession: S18030, S33570, A48332, S18029
R/Honda, M.; Kaneko, S.; Masashi, U.; Kodayashi, K.; Murakami, S.
Submitted to the EMBL Data Library, September 1991
A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated from a single patie
A/Reference number: S18028
A/Accession: S18030
A/Molecule type: genomic RNA
A/Residues: 1-3010 <HON>
A/Cross-references: EMBL:X61596; NID:959478; PIDN:CAA43793.1; PID:959479
A/Experimental source: isolate UK1 from an individual
R/Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A/Accession: S33570
A/Molecule type: genomic RNA
A/Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>
A/Cross-references: EMBL:X61591
A/Note: this sequence is inconsistent with the nucleotide translation
A/Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
as Trp, and TTC for residue 771 as Ser
A/Note: sequence extracted from NCBI database (NCBI:121747, NCBI:121748)
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EMP>
F/192-389/Product: major envelope protein E #status predicted <MEP>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-165/Product: nonstructural protein NS3 #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif A (P-loop)
F/1616-1862/Product: DEXH motif
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (AS

Query Match 93.2%; Score 1717; DB 1; Length 3010;
Best Local Similarity 92.3%; Pred. No. 3, 6e-136;
Matches 325; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

1 AHQWIPPLNVRGGRDAIILITCAVHPELIFDIYTKLLAIFGPMVLOAGITKVPYVR 60
855 ACQWVWPPPLNVRGGRDAIILITCAVHPELIFDIYTKLLAIFGPMVLOAGITKVPYVR 914
61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLDQMAHAGLDLAVAVEPV 120
915 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLDQMAHAGLDLAVAVEPV 974
121 IFSDMEVKITMGADTAACGDIISGLPVSAARGREILGPADNFEQGRLLAPITAYSQ 180
975 VFSDEMKITMGADTAACGDIISGLPVSAARGREILGPADNFEQGRLLAPITAYSQ 1034
181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSEFLATCVNGVCTVFEHAGSKTLAPX 240
1035 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSEFLATCVNGVCTVFEHAGSKTLAPX 1094
241 GPITQWYNNVQDVLVGMQAPPGASMTCTCGSSDLVLTREADVLPVRRGDSGSLIS 300
1095 GPITQWYNNVQDVLVGMQAPPGASMTCTCGSSDLVLTREADVLPVRRGDSGSLIS 1154
301 PRPVSYLKGSSGGPILCPSGHANGVIFRAAVCTRGAKAVDIPVESMETTMR 352
1155 PRPVSYLKGSSGGPILCPSGHANGVIFRAAVCTRGAKAVDIPVESMETTMR 1206

GNMWTG genome polyprotein - hepatitis C virus
N/Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
A/Variety: isolate UK1
C/Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text_change 19-Jan-2001
C/Accession: A38465
R/Takamizawa, A.; Mori, C.; Fukey, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J
J. Virol. 65, 1105-1113, 1991
A/Title: Structure and organization of the hepatitis C virus genome isolated from human c
A/Reference number: A38465; MUID:91140658; PMID:1847440
A/Accession: A38465
A/Molecule type: genomic RNA
A/Residues: 1-3010 <TKA>
A/Cross-references: EMBL:M58335; NID:9329770; PIDN:AAA72945.1; PID:9329771
A/Experimental source: isolate UK1 from an individual
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; capsid protein C #status predicted <CPC>
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EMP>
F/192-389/Product: major envelope protein E #status predicted <MEP>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-165/Product: nonstructural protein NS3 #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif A (P-loop)
F/1616-1862/Product: DEXH motif
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 92.2%; Score 1699; DB 1; Length 3010;
Best Local Similarity 91.5%; Pred. No. 1, 2e-134;
Matches 322; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

1 AHQWIPPLNVRGGRDAIILITCAVHPELIFDIYTKLLAIFGPMVLOAGITKVPYVR 60
855 ADLHWIPPLNVRGGRDAIILITCAVHPELIFDIYTKLLAIFGPMVLOAGITKVPYVR 914
61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLDQMAHAGLDLAVAVEPV 120
915 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTPLDQMAHAGLDLAVAVEPV 974
121 IFSDMEVKITMGADTAACGDIISGLPVSAARGREILGPADNFEQGRLLAPITAYSQ 180
975 VFSDEMKITMGADTAACGDIISGLPVSAARGREILGPADNFEQGRLLAPITAYSQ 1034
181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSEFLATCVNGVCTVFEHAGSKTLAPX 240
1035 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSEFLATCVNGVCTVFEHAGSKTLAPX 1094
241 GPITQWYNNVQDVLVGMQAPPGASMTCTCGSSDLVLTREADVLPVRRGDSGSLIS 300
1095 GPITQWYNNVQDVLVGMQAPPGASMTCTCGSSDLVLTREADVLPVRRGDSGSLIS 1154
301 PRPVSYLKGSSGGPILCPSGHANGVIFRAAVCTRGAKAVDIPVESMETTMR 352
1155 PRPVSYLKGSSGGPILCPSGHANGVIFRAAVCTRGAKAVDIPVESMETTMR 1206

RESULT 6

GNMWTG genome polyprotein - hepatitis C virus (strain HCV-1)
N/Contains: capsid protein C; envelope protein M; hepatitis B (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
A/Variety: isolate UK1
C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
C/Accession: A39166; PQ0403; PQ0404
R/Choo, Q.-J.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coit
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A/Title: Genetic organization and diversity of the hepatitis C virus.
A/Reference number: A39166; MUID:91172826; PMID:1848704
A/Accession: A39166

A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:G328973; PIDN:AAA45676.1; PID:G328974
R:Chan, S.W.; McComish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: PQ0393; MUID:92268871; PMID:1316939
A:Accession: PQ0403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CHA>
A:Cross-references: DDBJ:D10128
A:Experimental source: Isolates E-b16
A:Accession: PQ0404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: Isolates E-b17
C:Superfamily: hepatitis C virus genome polypeptide, glycoprotein, hydrolase; nonstructura
C:Keywords: ATP; capsid protein; envelope protein; #status predicted <CPC>
F:115-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-389/Product: major envelope protein E #status predicted <MEB>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 87.7%; Score 1616; DB 1; Length 3011;
Best Local Similarity 84.1%; Pred. No. 12e-127;
Matches 296; Conservative 28; Mismatches 28; Indels 0; Gaps 0;

QY 1 AHIQWIPPLNVRGGRDAIILLTCVAHPELFDITKLLAIFGPMVLQAGITKVPYFR 60
DB 855 AQLHWVPEPLNVRGGRDAVILLMCVAPLTFVFDITKLLAVGPEMLQSLKVPYFR 914
QY 61 AGLIRACMYRKAAGHYQMAFMKLAALTGYVVDHPLDPMANAGRLDAVAPEV 120
DB 915 VQGLRICALARKKQGHVQVMAIKLGAITGYVNHPLPLDMANAGRLDAVAPEV 974
QY 121 IFSDEMYKLTWGAADTAACDIIISGLPVSARKREIILGPADNEBQGMFLAPITAYQ 180
DB 975 VFSQMETKLTWGAADTAACDIIISGLPVSARKREIILGPADNEBQGMFLAPITAYQ 1034
QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 240
DB 1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 1094
QY 241 GPITQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRGDSRGLLS 300
DB 1095 GPVQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRGDSRGLLS 1154
QY 301 PRVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAXVDFIVESMETMR 352
DB 1155 PRPISYLKSSGGPILCPAGHAGVIFRAAVCTRGVAAXVDFIVESMETMR 1206

RESULT 7
S40770
genome polypeptide - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770

A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:G221586; PIDN:BA01582.1; PID:G221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Ugn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116; PMID:2170712
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:G221511; PIDN:BA00705.1; PID:G221512
A:Experimental source: Isolate HC-11
C:Superfamily: hepatitis C virus genome polypeptide, nucleotide binding; P-loop; polypeptide; serine
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
F:115-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-389/Product: major envelope protein E #status predicted <MEB>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 87.4%; Score 1610; DB 1; Length 3011;
Best Local Similarity 83.5%; Pred. No. 4e-127;
Matches 294; Conservative 28; Mismatches 30; Indels 0; Gaps 0;

QY 1 AHIQWIPPLNVRGGRDAIILLTCVAHPELFDITKLLAIFGPMVLQAGITKVPYFR 60
DB 855 AQLHWVPEPLNVRGGRDAVILLMCVAPLTFVFDITKLLAVGPEMLQSLKVPYFR 914
QY 61 AGLIRACMYRKAAGHYQMAFMKLAALTGYVVDHPLDPMANAGRLDAVAPEV 120
DB 915 VQGLRICALARKKQGHVQVMAIKLGAITGYVNHPLPLDMANAGRLDAVAPEV 974
QY 121 IFSDEMYKLTWGAADTAACDIIISGLPVSARKREIILGPADNEBQGMFLAPITAYQ 180
DB 975 VFSQMETKLTWGAADTAACDIIISGLPVSARKREIILGPADNEBQGMFLAPITAYQ 1034
QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 240
DB 1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCNVCMTVYHGAAGRTIASPK 1094
QY 241 GPITQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRGDSRGLLS 300
DB 1095 GPVQWNTNVDQLVGMQAPPGARSMTPTCGSSDLYLVRHADVIIVRRGDSRGLLS 1154
QY 301 PRVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAAXVDFIVESMETMR 352
DB 1155 PRPISYLKSSGGPILCPAGHAGVIFRAAVCTRGVAAXVDFIVESMETMR 1206

RESULT 8
G40770
genome polypeptide - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A36814; A41546
R:Inchausti, G.; Zebende, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: CC
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>

A/Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
 R:Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A/Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison
 A/Reference number: A41566; MUID:92052256; PMID:1658800
 A/Contents: annotation
 A/Note: neither amino acid nor nucleotide sequence is given
 A/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1.115/Product: capsid protein C #status predicted <GPC>
 F:1.116-191/Product: envelope protein M #status predicted <EPM>
 F:1.192-389/Product: major envelope protein E #status predicted <MEP>
 F:3.90-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:7.30-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1.007-1616/Product: hepatitis C virus status predicted <NS3>
 F:1.230-1237/Product: nucleotide-binding motif A (P-loop)
 F:1.317-1317/Region: nucleotide-binding motif B
 F:1.317-1317/Region: DEXH motif
 F:1.316-1862/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1.863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2.014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:1.96-2.09-2.34,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 87.1%; Score 1605; DB 1; Length 3011;
 Best Local Similarity 83.5%; Pred. No. 1e-126; Indels 0; Gaps 0;
 Matches 294; Conservative 30; Mismatches 28; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGDAIILLTCVAPBELIFDITKLLAIFGFLMWLAGITKVPYFVR 60
 DB 855 AQLHWVPLNVRGGDAVIMLTCTVHPALVFDTIKLLAIFGLMWLAGITKVPYFVR 914
 QY 61 AAGLIRACMLVKKAGHYQVMAFMKLAALTCTYVYDHLPLQDMHAGRLDAVAVEPV 120
 DB 915 VQGLRICALARKAGHYQVMAFMKLAALTCTYVYDHLPLQDMHAGRLDAVAVEPV 974
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNPEGGRLAPITAYSQ 180
 DB 975 VFSMEKLTITWGADTAACGDIISGLPVARSRGREILLGPADNPEGGRLAPITAYSQ 1034
 QY 181 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCWTFVHAGSKTLAEPK 240
 DB 1035 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOTFLATCINGVCWTFVHAGSKTLAEPK 1094
 QY 241 GPITQWTVNDODLVGMOPPGARSMPTCGSSDLVYTRHADYVPRRRGDSRSLLS 300
 DB 1095 GPVITQWTVNDODLVGMOPPGARSMPTCGSSDLVYTRHADYVPRRRGDSRSLLS 1154
 QY 301 PRPVSYLKSGSGGELLCPSGHNAVGFRAAVCTRGVAKAVDFIPVSMETMR 352
 DB 1155 PRPISYLKSGSGGELLCPSTGHNAVGFRAAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 9
 JC5620 genome polyprotein - hepatitis C virus (isolate EUH1480)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
 protein NS4; nonstructural protein NS4b; nonstructural protein NS5
 C/Species: hepatitis C virus
 C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C/Accession: J05620
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997
 A/Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
 A/Reference number: J05620; MUID:9736593; PMID:9223423
 A/Accession: J05620
 A/Molecule type: tRNA
 A/Residues: 1-3014 <CHA>
 A/Cross-references: GB:Y113184
 A/Experimental source: genotype 5a, which predominates in South Africa
 A/Note: the translation of the nucleotide sequence is not complete in this paper
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>

F:1.16-191/Product: envelope protein M #status predicted <EPM>
 F:1.92-389/Product: major envelope protein E #status predicted <MEP>
 F:1.84-408/Region: hypervariable #status predicted <MB>
 F:3.90-730/Product: nonstructural protein NS1 #status predicted <NS1>
 F:7.31-1007/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1.008-1616/Product: hepatitis C virus status predicted <NS3>
 F:1.231-1238/Region: nucleotide-binding motif A (P-loop)
 F:1.313-1318/Region: nucleotide-binding motif B
 F:1.317-1320/Region: DEXH motif
 F:1.317-1663/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1.864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2.015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2.210-2.249/Region: interferon sensitivity determining #status predicted

Query Match 76.1%; Score 1401; DB 1; Length 3014;
 Best Local Similarity 70.7%; Pred. No. 1.8e-109; Indels 0; Gaps 0;
 Matches 249; Conservative 51; Mismatches 52; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGDAIILLTCVAPBELIFDITKLLAIFGFLMWLAGITKVPYFVR 60
 DB 856 AQLHWVPLNVRGGDAVIMLTCTVHPALVFDTIKLLAIFGLMWLAGITKVPYFVR 915
 QY 61 AAGLIRACMLVKKAGHYQVMAFMKLAALTCTYVYDHLPLQDMHAGRLDAVAVEPV 120
 DB 916 AQLHWVPLNVRGGDAVIMLTCTVHPALVFDTIKLLAIFGLMWLAGITKVPYFVR 975
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNPEGGRLAPITAYSQ 180
 DB 976 VFSMEKLTITWGADTAACGDIISGLPVARSRGREILLGPADNPEGGRLAPITAYSQ 1035
 QY 181 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOSFLATCVNGVCWTFVHAGSKTLAEPK 240
 DB 1036 QTRGLGCIITSLTGRDKNOVEGEVQVSTATOTFLATCINGVCWTFVHAGSKTLAEPK 1095
 QY 241 GPITQWTVNDODLVGMOPPGARSMPTCGSSDLVYTRHADYVPRRRGDSRSLLS 300
 DB 1096 GPVITQWTVNDODLVGMOPPGARSMPTCGSSDLVYTRHADYVPRRRGDSRSLLS 1155
 QY 301 PRPVSYLKSGSGGELLCPSGHNAVGFRAAVCTRGVAKAVDFIPVSMETMR 352
 DB 1156 PRPISYLKSGSGGELLCPSTGHNAVGFRAAVCTRGVAKAVDFIPVSMETMR 1207

RESULT 10
 GNRVJ8 genome polyprotein - hepatitis C virus (strain HC-98)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
 protein NS4; nonstructural protein NS4b; nonstructural protein NS5
 C/Species: hepatitis C virus
 C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C/Accession: A40250; E00397; E00559
 R:Okamoto, H.; Kuri, K.; Okada, S.T.; Yamamoto, K.; Litzuka, H.; Tanaka, T.; Fukuda, S.;
 Virology 189, 331-341, 1992
 A/Title: Full-length sequence of a hepatitis C virus genome having poor homology to repo
 A/Reference number: A40250; MUID:9223032; PMID:1314459
 A/Accession: A40250
 A/Molecule type: genomic RNA
 A/Residues: 1-3033 <OKR>
 A/Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BA01761.1; PID:G221609
 R:Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Pauley, J.F.; Follett, E.; Yap, P.L.
 J. Gen. Virol. 73, 1131-1141, 1992
 A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to ex
 A/Reference number: P00393; MUID:92268871; PMID:1316939
 A/Accession: P00393
 A/Molecule type: genomic RNA
 A/Residues: 2678-2754 <CHA>
 A/Cross-references: DDBJ:D10134
 A/Experimental source: isolate E-b12
 R:Kato, N.; Ootsuyama, Y.; Okochi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991
 A/Title: Distribution of plural HCV types in Japan.
 A/Reference number: P00554; MUID:92068204; PMID:11720309
 A/Accession: P00554

QY 239 PKGPIITOMYNTVDDIVGQWQAPPGARSMPTCTCGSSDLVYTRADYIPVRKRGDSRSL 238
 Db 1044 PVGVNCRWNSPSPDVAVPPPLPSGASCLEPCCKGTGQWICRN--DALCHGRISKLVEL 1101
 QY 299 LSPRPVSLKSGSGPLLCPSGHAIVGIFRAAVCTRGV-----AKAVDFPVES 346
 Db 1102 DLPFEISDFRSGSSPLICDEGHVGMV--VSVLHRGVKVTGVKRWETLTPKDS 1155

RESULT 13

108839
 polyprotein - marmoset hepatitis GB virus A
 C:Species: marmoset hepatitis GB virus A
 C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
 C:Accession: T08839
 R:Ekker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushaiwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A:Reference number: Z16486; MUID:98120818; PMID:9460920
 A:Accession: T08839
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: genomic RNA
 A:Residues: 1-2970 <ERR>
 A:Cross-references: EMBL:AF023424; NID:g2828597; PIDN:AC40501.1; PID:g2828598
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: polyprotein

Query Match 18.6%; Score 342.5; DB 2; Length 2970;
 Best Local Similarity 28.9%; Pred. No. 4.2e-20;
 Matches 103; Conservative 56; Mismatches 133; Indels 65; Gaps 12;

QY 37 LLLAIFGP-----LMVLQAGITKYPYFAQGLIRACMLVKKXAGHYVMAFMKA 88
 Db 816 LVAMFMPREIAAVCAVFIIGFPPVVDILFVLLVSSPNVRLARVLDLSVAADKXA 875
 QY 89 ALTGIVYDHLTLPD--WAAG-----LRDLAVERPVSFSDMEVKITWGADTA 137
 Db 876 T--TWLVEKLRKNCFLVMAQVTRTAELRQMGFLPEVAVAPEDCAMVRDARFL 932
 QY 138 ACGDIIISGLPVSARRGRREILLGPADNFEQGWRL-----LAPITXSSQITGLGCIIT 191
 Db 933 SCQGSYHGRVYVARRDEVLIGLVN---WELPFGFVPTAAVVA--HGGGFGFVKT 987
 QY 192 SLTGKXNOVEGEVQVYVSTATQSFATCVNGVCMVFHSGASKTLAIPKPIITOMYNTD 251
 Db 988 SMTGMDTEHVGVVAVLGTSTSMGTGVNGVYTYHGSNAKTLAQMGPVNSRWMSAS 1047
 QY 252 ODLVGMQAPPGARSMPTCTCGSSDLVYTRADYIPVRKRGDSRGLS-----300
 Db 1048 DVAIVAPLVGAKCEPCCKQPOGVVVI-----RND--GALCHGILGRVVELDL 1094
 QY 301 PRPVSLKSGSGPLLCPSGHAIVGIFRAAVCTRG-----VAKAVDFPVESMETT 350
 Db 1095 PAELCDFRSGSSGPLICDEGHAVGM--ISVLHRGSVVTGIRTKPWEITLPRALITHT 1150

RESULT 14

VHMMH2
 structural protein 2 precursor - hepatitis E virus (strain Burma)
 C:Species: hepatitis E virus
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Jul-1999
 C:Accession: C40778
 R:Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.
 Virology 185, 120-131, 1991
 A:Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi
 A:Reference number: A40778; MUID:92024067; PMID:1926770
 A:Accession: C40778
 A:Molecule type: genomic RNA
 A:Residues: 1-660 <TAM>
 A:Cross-references: GB:M73218; NID:9330023; PIDN:AAA45736.1; PID:g330026
 A:Note: the authors translated the codon GGC for residue 2 as Ala
 C:Superfamily: hepatitis E virus structural protein 2

C:Keywords: structural protein
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-660/Product: structural protein 2 #status predicted <SP2>

Query Match 5.6%; Score 102.5; DB 1; Length 660;
 Best Local Similarity 19.8%; Pred. No. 1.2;
 Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

QY 35 TKLL--ALRGPVNLVQAG-----ITKPYEVR--AQGLIRACMLVKKXAGHYVQWA 83
 Db 151 TNLVLYAAPLSPLPLQDGTNTHIMATBSNVAQYRVARATIRYRPLVNNAGVAIGIS 210
 QY 84 EKKALATCTYYVDHITPLQPMHAGLADLAIVAPVPSFDEVKIITWGADTAACGDI 143
 Db 211 FWPQTTTPTSV-----DNMSITSDVRLVQPGIASBELVI-----246
 QY 144 SGIPVSARRGRREILLGPAD--NFEQGWRLAPI--TAVSQOTRGL-----GCITSLTG 195
 Db 247 -----PSERLHYRNQCMRSVTSVGAEEATSGVMCTHGSIVNSYTN 290
 QY 196 -----RDKNOYGEVQVYVSTATQSF 216
 Db 291 TPTGALGLDPLALEFRNLTPGNTNTRVRSSTARRLRGADGAEITTTATATFM 350
 QY 217 A-----TCVNGV-----CMYVEH-----GAG-----SKTLAPRG--PIT 244
 Db 351 KDLVTSINVGELRGALILFNLADTLGLLPTELISAGGULFYSPVVSANGEPV 410
 QY 245 QMYNTVDDIVGQWQAPPGARSMPTCTCGSSDLVY---TRADYIPVRKRGDSRG--SLIS 300
 Db 411 KLYTSVENA-----QDDKGIATPHDIDGESRVVLDYDNGHQRPTSPAPSRFSVLR 466
 QY 301 PRPVSLK-----GSSGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIV 344
 Db 467 ANDVLMISLTAEXYDQSTYSGSTGPVYV--SDSVTLVNVVATGAQAVARSLDMTKY 519

RESULT 15

H71426
 hypothetical protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: Columbia
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
 C:Accession: H71426
 R:Bayan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
 P.; Wedler, H.; Wedler, E.; Wandutt, R.; Weitzenecker, T.; Pohl, T.M.; Terryn, N.; Giehl
 vanagh, T.; Hempel, S.; Kotter, P.; Shtilian, K.D.; Rieger, M.; Schaeffer, W.; Funk, B.
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,
 ehoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reehman, S.; Anic
 C.; Chaitatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
 A:Reference number: A71400; MUID:98121113; PMID:9461215
 A:Accession: H71426
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-692 <BEV>
 A:Cross-references: GB:Z97340; NID:g2244950; PID:e327492; PID:g2244965
 C:Genetics:
 A:Map position: 4C09-4G3845

Query Match 5.6%; Score 102.5; DB 2; Length 692;
 Best Local Similarity 23.9%; Pred. No. 1.2;
 Matches 91; Conservative 41; Mismatches 127; Indels 121; Gaps 22;

QY 8 PPLVNRGGRDAIITLCVAPHEIIPDITKLLAFGIMVTLQAGITVPP--TVRAGGL 65
 Db 101 PTVASVAGSNICIE---KPEIIMDTIELLR--FLTLQWL-----FSKKPVAVLESNGYT 151
 QY 66 PACMLVR--RAAGHYVQWAFM-----KLAALGTGYVDHLTP 101
 Db 152 HEDVTLQPKRAGVGHIMQAPAFITIRDTNSKILLIRGTHSINKDTLTAAGAVVPPFHHSV 211

```
QY 102 LODWAHAGLRLDPLAVAVEPVPISPMDEVKIITWGADTAACGDIISGLPYSARGREILLGPA 161
| | | | |
DB 212 LHD---GGLSNLVVIGY-----AHG-----WVAAAR----- 234
| | | | |
QY 162 DNFEGQW--RLAPITAYSOQTRGLLCIITSLTGRDKNOVEGEVQVSTA---TQSF 215
| | | | |
DB 235 -----WIAKLSVP-----CLKAL---DENP-SRKVOIVGHSILGGGTASL 270
| | | | |
QY 216 LATCV-----NGVCWTFVHAGSKTLAGPKP--ITOMVTWVDDOLV---GMOAPGA 263
| | | | |
DB 271 LTYILREKXERASATCTFAPGTENMINESGKHFTITTING--SDLVPTFSASSVDL 328
| | | | |
QY 264 RSMTPCTCGSSDLYLVTRHADVIP-VRRGDSRGSLLSPRPVSYLKSSGGPULCP--SG 320
| | | | |
DB 329 RSEVTSSMSNDLRDQVHTRVLSVYRSATAIIGSRLLPISAKAKAVAGAGAILIRPVSSG 388
| | | | |
QY 321 HAVGIFRAAVCTRGVAKAVD 340
| | | | |
DB 389 TQVAAFLVNGC--GKIKCID 406
| | | | |
```

Search completed: May 6, 2004, 09:37:19
Job time : 12.4639 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 7.34864 Seconds

(without alignments)
2494.160 Million cell updates/sec

Title: US-10-650-585-13
Perfect score: 1842

Sequence: 1 AHQVWIPINVRGRDAII.....RGVAKAVDFIPVESMTTNR 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1773	96.3	3010	1	P26662 h genome po
2	1766	95.9	3010	1	POLG_HCVUA
3	1749	95.0	3010	1	POLG_HCVUT
4	1699	92.2	3010	1	POLG_HCVTW
5	1616	87.7	3011	1	POLG_HCVBK
6	1605	87.1	3011	1	POLG_HCVI
7	1303	70.7	3033	1	POLG_HCVH
8	1301	70.6	3033	1	POLG_HCVU8
9	1202.5	5.6	660	1	POLG_HCVU6
10	102.5	5.6	660	1	VST2_HEVPA
11	101	5.5	564	1	SRSOC_AKATH
12	101	5.5	600	1	DPO2_MOUSE
13	93.5	5.1	1380	1	CYAA_LEIDO
14	93	5.0	434	1	TOLB_CHLTE
15	92.5	5.0	706	1	TRFE_HORSE
16	92	5.0	659	1	VST2_HEYME
17	91.5	5.0	485	1	VST2_HEYRH
18	91.5	5.0	660	1	VST2_HEYVY
19	91.5	5.0	3414	1	POLG_TBEVW
20	90.5	4.9	3414	1	POLG_TBEVH
21	89.5	4.9	961	1	ATCU_YERPE
22	87.5	4.8	3412	1	POLG_TBEVS
23	86.5	4.7	347	1	MDHM_EUCGU
24	86.5	4.7	470	1	NRAM_IAMHM
25	86	4.7	338	1	GALB_NERGO
26	86	4.7	433	1	DCUA_WOLSU
27	85	4.6	470	1	NRAM_IATRA
28	85	4.6	730	1	HELS_METMA
29	85	4.6	854	1	PWP2_SCHPO
30	85	4.6	3313	1	CLR3_RAT
31	84.5	4.6	1705	1	PTPV_MOUSE
32	84	4.6	309	1	UCP2_RAT
33	84	4.6	339	1	GPDA_COREF

34	84	4.6	423	1	MK09_MOUSE	Q9WTU6 mus musculi
35	84	4.6	470	1	NRAM_IARDE	P31510 Influenza a
36	84	4.6	503	1	YA02_MYCTU	O05586 mycobacteri
37	83.5	4.5	538	1	DAC_ACRSP	P39045 actinomadar
38	83.5	4.5	1399	1	RPOC_PSEAE	Q9HVC9 pseudomonas
39	83	4.5	341	1	MDHM_BRANA	Q43744 brassica na
40	82.5	4.5	453	1	NRAM_IAMIT	P03470 influenza a
41	82.5	4.5	711	1	H2PA_RHIME	Q8VT23 rhizobium m
42	82.5	4.5	1022	1	CA26_CHICK	F15988 gallus gall
43	82	4.5	309	1	UCP2_MOUSE	P70406 mus musculi
44	82	4.5	384	1	MK08_BRARE	Q9GDS4 brachydanio
45	82	4.5	403	1	PGK_CHILMU	Q9PIN4 chlamydia m

ALIGNMENTS

```

RESULT 1
ID POLG_HCVUA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (N21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (Isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijioka M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijioka M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL, D90208; BAA14233.1; -

```

DR PIR; A39253; GNMVCT.
 DR HSPB; P26663; LUXP.
 DR MEROS; 829.001; -.
 DR INTERPRO; IPR009003; Cys_Ser_lypsin.
 DR INTERPRO; IPR001410; DEAD.
 DR INTERPRO; IPR002522; HCV_capsid.
 DR INTERPRO; IPR002521; HCV_core.
 DR INTERPRO; IPR002519; HCV_env.
 DR INTERPRO; IPR002518; HCV_NS1.
 DR INTERPRO; IPR002518; HCV_NS2.
 DR INTERPRO; IPR000745; HCV_NS4a.
 DR INTERPRO; IPR001490; HCV_NS4b.
 DR INTERPRO; IPR002868; HCV_NS5a.
 DR INTERPRO; IPR002166; HCV_RdRp.
 DR INTERPRO; IPR001650; Helicase_C.
 DR INTERPRO; IPR004109; Peptidase_C29.
 DR INTERPRO; IPR007095; RNA_pol_D5_PS.
 DR INTERPRO; IPR007094; RNA_pol_PSVlr.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR Pfam; PF016062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolyase; Serine protease.
 INT_MET 1
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1615
 CHAIN 1616 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 TRANSMEM 347 369
 ACT_SITE 1107 1107
 ACT_SITE 1165 1165
 NP_BIND 1230 1237
 SITE 1316 1319
 CARBOHYD 196 196
 CARBOHYD 209 209
 CARBOHYD 234 234
 CARBOHYD 250 250
 CARBOHYD 305 305
 CARBOHYD 417 417
 CARBOHYD 423 423
 CARBOHYD 430 430
 CARBOHYD 448 448
 CARBOHYD 532 532
 CARBOHYD 556 556
 CARBOHYD 576 576
 CARBOHYD 623 623
 CARBOHYD 645 645
 CARBOHYD 2041 2041
 CARBOHYD 2077 2077
 CARBOHYD 2240 2240
 CARBOHYD 2788 2788
 SEQUENCE 3010 AA; 327017 MW; AA937944F6DB185 CRC64;

Query Match 96.3%; Score 1773; DB 1; Length 3010;

Best Local Similarity 94.6%; Pred. No. 3e-141.
 Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHLOWIPLNVRGGRDAIILLTCVAPPELIIFDTIKLLAIFGLMVLQAGITKVPYFVR 60
 DB 855 AHLOWVPLNVRGRDAIILLTCVAPPELIIFDTIKLLAIFGLMVLQAGITKVPYFVR 914
 QY 61 AAGLIACMLVRKAAGHYVOMAMKLAALTGTYYVHLLPLQDMANAG;RDLAAVAEPV 120
 DB 915 AAGLIACMLVRKAAGHYVOMAMKLAALTGTYYVHLLPLRLWALAG;RDLAAVAEPV 974
 QY 121 IFSDMEVKIITWAGDTAACGDIISGLPYVSARGSEIILGPADEGQWLLAPITAYSQ 180
 DB 975 VFSDEMETLITWAGDTAACGDIISGLPYVSARGSEIILGPADEGQWLLAPITAYSQ 1034
 QY 181 QTRGLGCIITSLTGRDNQVYEGVQVSTATQSFATCTNGVCWTYFHAGSKITLAPK 240
 DB 1035 QTRGLGCIITSLTGRDNQVYEGVQVSTATQSFATCTNGVCWTYFHAGSKITLAPK 1094
 QY 241 GPITQMTNNVDLVGMQAPRGASMTPTCGSSDLYLVRHADVIFVRSGDSRGLLS 300
 DB 1095 GPITQMTNNVDLVGMQAPRGASMTPTCGSSDLYLVRHADVIFVRSGDSRGLLS 1154
 QY 301 PRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAVAFPIPVSEMTTMR 352
 DB 1155 PRPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGAVAFPIPVSEMTTMR 1206

RESULT 2
 ID POLG HCVJT STANDARD; PRT; 3010 AA.
 AC 000269;
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Genome polyprotein [contains: Capsid protein C (core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepaticin)
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate HC-JT) (HCV).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCB1_TaxID=31642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=118627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.,
 RT Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RT Virus Res. 23:39-53(1992).
 RL
 RL -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 RL hydrophobic, suggesting a possible membrane-related function. NS3
 RL and NS5 may play a role in the viral RNA replication.
 RL -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 RL precursor polyprotein, commonly with Asp or Glu in the P6
 RL position, Cys or Thr in P1 and Ser or Ala in P1'.
 RL -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 RL {RNA} (N).
 RL -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 RL lipoprotein envelope. The envelope consists of two proteins of
 RL protein M and glycoprotein E. The nucleocapsid is a complex of
 RL protein C and RNA.
 RL -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ibb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL, D11168; EAA01943.1; -

DR PIR: A45573; A45573.

DR MEROPS; S29.001; -

DR MEROPS; U39.001; -

DR InterPro; IPR003003; Cys_Ser_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR003522; HCV_Capsid.

DR InterPro; IPR005211; HCV_Core.

DR InterPro; IPR002519; HCV Env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRp.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; Peptidase_C29.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PStir.

DR Pfam; PF01543; HCV_Capsid; 1.

DR Pfam; PF01542; HCV_Core; 1.

DR Pfam; PF01539; HCV Env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; Helicase_C; 1.

DR Pfam; PF00998; Viral_RdRp; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural

KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural

FT INIT_MET 1

FT CHAIN 1 115

FT CHAIN 192 191

FT CHAIN 384 383

FT CHAIN 730 729

FT CHAIN 1007 1006

FT CHAIN 1616 1615

FT CHAIN 1863 1862

FT CHAIN 2014 2013

FT CHAIN 347 369

FT TRANSMEM 1083 1083

FT ACT_SITE 1107 1107

FT ACT_SITE 1165 1165

FT NP_BIND 1230 1237

FT SITE 1316 1319

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 234 234

FT CARBOHYD 250 250

FT CARBOHYD 305 305

FT CARBOHYD 417 417

FT CARBOHYD 423 423

FT CARBOHYD 430 430

FT CARBOHYD 448 448

FT CARBOHYD 532 532

FT CARBOHYD 540 540

FT CARBOHYD 556 556

FT CARBOHYD 576 576

FT CARBOHYD 623 623

FT CARBOHYD 645 645

FT CARBOHYD 2041 2041

FT CARBOHYD 2077 2077

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3010 AA; 326573 MW; 94A1C7435D642DB CRC64;

Query Match 95.9%; Score 1766; DB 1; Length 3010;

Best Local Similarity 94.9%; Pred. No. 1,2e-140;

Matches 334; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 AHQVWIPPLNVRGGDAIILLTCVAVPELIDITKLLAIFGLPMVTOAGITKPYFVR 60

DB 855 AHLQVWIPPLNVRGGDAIILLTCVAVPELIDITKLLAIFGLPMVTOAGITKPYFVR 914

QY 61 AAGLIACGLVAKKAGGHVQMAFMKLAALCTGVYDHTLPQDNHAGLRLAVAEV 120

DB 915 AAGLIACGLVAKKAGGHVQMAFMKLAALCTGVYDHTLPQDNHAGLRLAVAEV 974

QY 121 IFSDEMKIITWGADTAACGDIISGLPVSRGRREILLGPADNFGSGWRLLAPITAYSQ 180

DB 975 VFSDEMKIITWGADTAACGDIISGLPVSRGRREILLGPADNFGSGWRLLAPITAYSQ 1034

QY 181 QTRGLIGCIITSLTGRDKQVGEVQVYSTAQSLATCVAGVCTVPHGASKTLAPK 240

DB 1035 QTRGLIGCIITSLTGRDKQVGEVQVYSTAQSLATCVAGVCTVPHGASKTLAPK 1094

QY 241 GPITQWYTVDDDLVGMQAPGARSMTPTCGSSDLYLVTRADYIPVRRRDSGSLIS 300

DB 1095 GPITQWYTVDDDLVGMQAPGARSMTPTCGSSDLYLVTRADYIPVRRRDSGSLIS 1154

QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETMR 352

DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETMR 1206

RESULT 3

ID POLG_HCVTW STANDARD; PRT; 3010 AA.

AC P29846;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

DE Hepatitis C virus [isolate Taiwan] (HCV).

OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OC NCBI_TaxID=31645;

OX NCBI_TaxID=31645;

RN SEQUENCE FROM N.A.

RP MEDLINE:92230206; PubMed:1314449;

RX Chen P.-J., Lin M.-H., Tai K.-F., Liu P.-C., Lin C.-J., Chen D.-S.; "The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' terminal of viral genomic and antigenomic RNA."; Virology 188:102-113 (1992).

-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 catalytic activity: Hydrolysis of four peptide bonds in the viral position, Cys or Thr in P1 and Ser or Ala in P1'.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).

-1- SUBUNIT: The viron of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL, M84754; -, NOT_ANNOTATED_CDS.

PIR, A40244; GKMVTV.

PDB, 1N64; 25-FEB-03.

PDB, 1NS3; 08-APR-98.

MEROPS, S29.001; -.

MEROPS, U39.001; -.

InterPro, IPR009003; Cys_Ser_trypsin.

InterPro, IPR001410; DEAD.

InterPro, IPR002522; HCV_capsid.

InterPro, IPR002521; HCV_core.

InterPro, IPR002519; HCV_env.

InterPro, IPR002511; HCV_NSI.

InterPro, IPR002518; HCV_NS2.

InterPro, IPR002515; HCV_NS4.

InterPro, IPR001490; HCV_NS4B.

InterPro, IPR002868; HCV_NS8.

InterPro, IPR002166; HCV_RdRp.

InterPro, IPR001650; Helicase_C.

InterPro, IPR004109; Peptidase_C29.

InterPro, IPR007095; RNA_pol_DS_PS.

InterPro, IPR007094; RNA_pol_PSVlr.

Pfam, PF01543; HCV_capsid; 1.

Pfam, PF01542; HCV_core; 1.

Pfam, PF01539; HCV_env; 1.

Pfam, PF01560; HCV_NSI; 1.

Pfam, PF01538; HCV_NS2; 1.

Pfam, PF02907; HCV_NS3; 1.

Pfam, PF01006; HCV_NS4; 1.

Pfam, PF01001; HCV_NS4B; 1.

Pfam, PF01506; HCV_NS8; 1.

Pfam, PF00271; Helicase_C; 1.

Pfam, PF00998; Viral_RdRp; 1.

ProDom, PD186062; HCV_NSI; 1.

SMART, SMO0487; DEXDC; 1.

Polypeptide, Glycoprotein; Transferrin; RNA-directed RNA polymerase;

Core protein; Coat protein; Envelope protein; Helicase; Arg-binding;

Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;

3D-structure.

INIT_MET 1 1

CHAIN 1 115

CHAIN 116 191

CHAIN 192 383

CHAIN 384 729

CHAIN 730 1006

CHAIN 1007 1615

CHAIN 1616 1862

CHAIN 1863 2013

CHAIN 2014 3010

CHAIN 3011 369

CHAIN 370 1083

CHAIN 1084 1107

CHAIN 1108 1165

CHAIN 1166 1237

CHAIN 1238 1319

CHAIN 1320 1369

CHAIN 1370 209

CHAIN 210 233

CHAIN 234 250

CHAIN 251 305

CHAIN 306 417

CHAIN 418 423

CHAIN 424 430

CHAIN 431 448

CHAIN 449 448

CHAIN 449 448

CHAIN 449 448

CHAIN 449 448

CHAIN 449 448

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDE215 CRC64;

Query Match 95.0%; Score 1749; DB 1; Length 3010;
Best Local Similarity 92.9%; Pred. No. 3.2e-139; Indels 0; Gaps 0;
Matches 327; Conservative 14; Mismatches 11;

QY 1 AHLQWIPPLNVRGGRDAIILLTCVAPBELIFDTKLLAIIFSPVLYLQAGITKPYFVR 60
DB 855 AHLQWIPPLNVRGGRDAIILLTCVAPBELIFDTKLLAIIFSPVLYLQAGITKPYFVR 914
QY 61 AGLIRACMLVRAAGHYQYQAFMKLALTCVYVYDHLPLQDMAHAGLRDAVAEPV 120
DB 915 AGLIRACMLVRAAGHYQYQAFMKLALTCVYVYDHLPLQDMAHAGLRDAVAEPV 974
QY 121 IFSDMEVYKILTWGADTAACGDIISGLPVARSREILGPADNFEQGWRLAPITAYSQ 180
DB 975 VFSDEMTKITWGDITACGDIILGLVVSARSGEILGPADNFEQGWRLAPITAYSQ 1034
QY 181 QTRGLLGCTITSLGRDNQYVEGVQVSTRTQSFALTQNGVQWTFPHAGSTTAGPK 240
DB 1035 QTRGLFGCIITSLGRDNQYVEGVQVSTRTQSFALTQNGVQWTFPHAGSTTAGPK 1094
QY 241 GPITQMTNVNODLVGMQAPGARSMTPCTGSSDLYLVRHADVIFVRRGDSRGLLS 300
DB 1095 GPITQMTNVNODLVGMQAPGARSMTPCTGSSDLYLVRHADVIFVRRGDSRGLLS 1154
QY 301 PRPVSYLNGSSGGLPLCPGSHAVGIFRAAVCTRGAAVDFIYVSEMTTWK 352
DB 1155 PRPVSYLNGSSGGLPLCPGSHAVGIFRAAVCTRGAAVDFIYVSEMTTWK 1206

RESULT 4
ID POLG HCVRK STANDARD; PRT; 3010 AA.
AC P26663;
DT 01-AUG-1992 (Rel. 23, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitisvirus.
OX NCBI_Taxid=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takamiizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Anoch T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers";
RL J. Virol. 65:1105-1113 (1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kornetceky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits


```

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCB5AF9 CRC64;

```

```

Query Match 87.7%; Score 1616; DB 1; Length 3011;
Best Local Similarity 84.1%; Pred. No. 5,5e-128;
Matches 296; Conservative 28; Mismatches 28; Indels 0; Gaps 0;

```

```

QY 1 ARLQWVTPPLNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLPLVLAQGITKVPYFVR 60
DB 855 AQLHWVTPPLNVRGGRDAIILLTCVAHPELIFDITKLLAIFGLPLVLAQGITKVPYFVR 914
QY 61 AGLIIRACMEVRYKAGGHHVQVAFMKLAALTGTVVYDHLPLDQMAHAGRLDAVAPEV 120
DB 915 VQGLRFLCARLARKVIGSHVQVMTIKGALITGVYVHLLPLDQMAHAGRLDAVAPEV 974
QY 121 IFSMEVKIITWGAADTAACDITISGLFVSARGRSEILLGPADNFEQGMWLLAPITAYSQ 180
DB 975 VFSQMEKLTITWGAADTAACDITISGLFVSARGRSEILLGPADNFEQGMWLLAPITAYSQ 1034
QY 181 QTRRLCCITITSLTGRKNQVGEVQVSTATQSFATCVNGCVTFYFHASGSKITLAGPK 240
DB 1035 QTRRLCCITITSLTGRKNQVGEVQVSTATQSFATCVNGCVTFYFHASGSKITLAGPK 1094
QY 241 GPIPTMTYDQDLVGMQAPGARSMTPCGSSDLVLYRHADVIPVRRCDSRGSILS 300
DB 1095 GPIPTMTYDQDLVGMQAPGARSMTPCGSSDLVLYRHADVIPVRRCDSRGSILS 1154
QY 301 PRPVSYIKGSSGGPLCPSGHAGVIFPAVCTRCVAVADFIPEHSETTMR 352
DB 1155 PRPVSYIKGSSGGPLCPSGHAGVIFPAVCTRCVAVADFIPEHSETTMR 1206

```

```

RESULT 6
POLG_HCVH STANDARD; PRT; 3011 AA.
ID POLG_HCVH
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP33) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)}.
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchausti G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RL virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
[2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

```

```

RX MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX MEDLINE=96154321; PubMed=9493270;
RA Kim U.U., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M67463; AAA5534.1; -.
DR PIR; A36814; GNVVCH.
DR PDB; 1HEI; 25-NOV-98.
DR PDB; 1AIV; 16-FEB-99.
DR PDB; 1AIR; 17-JUN-98.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR TRANSFAC; T04155; -.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002511; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4A.
DR InterPro; IPR001490; HCV NS4B.
DR InterPro; IPR002868; HCV NS5A.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_C29.
DR InterPro; IPR007094; RNA_pol_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01558; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.

```


CC	HEPACIVIRUS.
OX	NCBI_TaxID=11115,
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=92230232; PubMed=1314459;
RA	Okamoto H., Kura K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA	Fukuda S., Tada F., Mishiro S.;
RT	"Full-length sequence of a hepatitis C virus genome having poor
RT	homology to reported isolates: comparative study of four distinct
RT	genotypes.";
RL	Virology 188:331-341(1992).
CC	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC	hydrophobic, suggesting a possible membrane-related function. NS3
CC	and NS5 may play a role in the viral RNA replication.
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC	precursor polypeptide, commonly with Asp or Glu in the P6
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC	{RNA} (N).
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC	lipoprotein envelope. The envelope consists of two proteins:
CC	protein M and glycoprotein E. The nucleocapsid is a complex of
CC	protein C and RNA.
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC	-----
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its use
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC	or send an email to license@isb-sdb.ch).
CC	-----
DR	EMBL; D10388; BAA01761.1; -.
DR	PIR; A40250; GNAVYJ8.
DR	HSSP; P27958; 1HE1.
DR	MEROPS; 529.001; -.
DR	MEROPS; 529.001; -.
DR	InterPro; IPR0019003; Cys_Ser_trypsin.
DR	InterPro; IPR001410; DEAD.
DR	InterPro; IPR002522; HCV_capsid.
DR	InterPro; IPR002521; HCV_core.
DR	InterPro; IPR002519; HCV_env.
DR	InterPro; IPR002531; HCV_NS1.
DR	InterPro; IPR002518; HCV_NS2.
DR	InterPro; IPR000745; HCV_NS4a.
DR	InterPro; IPR001490; HCV_NS4b.
DR	InterPro; IPR002868; HCV_NS5A.
DR	InterPro; IPR002166; HCV_RDRP.
DR	InterPro; IPR004109; Peptidase_C29.
DR	InterPro; IPR007095; RNA_pol_DS_PS.
DR	InterPro; IPR007094; RNA_pol_PsVtr.
DR	Pfam; PF01543; HCV_capsid; 1.
DR	Pfam; PF01542; HCV_core; 1.
DR	Pfam; PF01539; HCV_env; 1.
DR	Pfam; PF01560; HCV_NS1; 1.
DR	Pfam; PF01538; HCV_NS2; 1.
DR	Pfam; PF02907; HCV_NS3; 1.
DR	Pfam; PF01006; HCV_NS4a; 1.
DR	Pfam; PF01001; HCV_NS4b; 1.
DR	Pfam; PF01506; HCV_NS5a; 1.
DR	Pfam; PF00998; Viral_RDRP; 1.
DR	ProDom; PD086062; HCV_NS1; 1.
DR	SMART; SM00487; DEXDC; 1.
KM	Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM	Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KM	protein; Hydrolase; Serine protease.
KM	Membrane; Nonstructural
FT	INIT_MET 1
FT	CHAIN 1 115
FT	CHAIN 116 191
FT	CHAIN 112 383
FT	CHAIN 384 733
FT	CHAIN
FT	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT	NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).

Query Match	Best Local Similarity	70.7%;	Score 1303;	DB 1;	Length 3033;
Matches 231;	Conservative 56;	Mismatches 65;	Indels 0;	Gaps 0;	
FT CHAIN	734	1010	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).		
FT CHAIN	1011	1619	PROTEASE/HELICASE NS3 (POTENTIAL).		
FT CHAIN	1620	1866	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).		
FT CHAIN	1867	2017	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).		
FT CHAIN	2018	3033	RNA-DIRECTED RNA POLYMERASE (POTENTIAL).		
FT TRANSMEM	347	369	POTENTIAL.		
FT ACT_SITE	1087	1087	CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT ACT_SITE	1111	1111	CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT ACT_SITE	1169	1169	CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT NP_BIND	1234	1241	ATP (POTENTIAL).		
FT SITE	1320	1323	DECH BOX.		
FT CARBOHYD	196	196	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	209	209	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	233	233	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	299	299	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	305	305	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	417	417	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	423	423	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	430	430	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	448	448	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	477	477	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	534	534	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	542	542	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	558	558	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	578	578	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	627	627	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	649	649	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	1091	1091	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	2038	2038	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	2359	2359	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT CARBOHYD	2811	2811	N-LINKED (GLCNAC. . .)	(POTENTIAL).	
FT SEQUENCE	3033 AA;	350177 MM;	1A173E7E9381E1D1A CRE64;		

01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=31767;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tem A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RA Fry K.E., Reyes G.R., (HEV): molecular cloning and sequencing of the
RT full-length viral genome.
RL Virology 185:120-131(1991).
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M73218; AAA45736.1; -
DR PIR; C40778; VHWMH2.
DR InterPro; IPR004261; SP2.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF03014; SP2; 1.
KW Signal.
FT CHAIN 1 19 POTENTIAL.
FT SIGNAL 20 660 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCCA461C CRC64;

Query Match 5.6%; Score 102.5; DB 1; Length 660;
Best Local Similarity 19.8%; Pred. No. 0.59; Indels 151; Gaps 19;
Matches 82; Conservative 52; Mismatches 130;
QY 35 TKLL--AIFGLMVLQAG-----ITKPYFVR--AQGLIRACMLYKAGGHVQWA 83
DB 151 TNLVLAAPLSPLPLQDGTNTHIMATEASNAQVRARATIRYRPLVNAVGTAISIS 210
QY 84 FPKLAALGTGYVYHLPLODMAHAGLDLAVAEVPIPSMEVKIITWADTAACGDI 143
DB 211 FWPQTITTSV-----DMNSITSDVRILVPGIASLVI----- 246
QY 144 SGLPVSARRGSEILLGPD--NFEQGWRLLAPI--TAVSOQTRGL-----GCITISLNG 195
DB 247 -----PSERLHYRNQGRSVETSGVAEEETSLVWLCHGSLVNSYTN 290
QY 196 -----RDXQVBEVQVVSATQSEFL 216
DB 291 TPYTGALGLDPALEFPRLTPGNTNTRYRSSTARHRLRGADGTALTTTAATRM 350
QY 217 A-----TCVNGV-----CMTVFH-----GAG-----SKTLAEPKG--PIT 244
DB 351 KDLYFTSTNGVGEIGRGIALTLFNLADTLGLPTELISSAGGLFYSRPVVSANGEPV 410
QY 245 QMTYTNVDQLVGMQAPGARSMTPCTCGSSDLVYV---TRHADVIYVRRRGDSRG--SLIS 300
DB 411 KLYTSVENA-----QDQKGIAPHDIDLGESRVVLDQYDNQHDQRPSPAPSRPFSVL 466
QY 301 PRPVSYLK-----GSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIV 344
DB 467 ANDVLMLSLTAAEYDQSTYGSSTGPFYV--SDSVTLVNVATGAQAVARSLDWTYV 519

01-FEB-1994 (Rel. 28, Created)
DE 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor (ORF2).
OS Hepatitis E virus (strain Pakistan) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=33774;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
RA Malik I.A., Iqbal M., Purcell R.H., (HEV): "Characterization of a prototype strain of hepatitis E virus."
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M80581; AAA5727.1; -
DR InterPro; IPR004261; SP2.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF03014; SP2; 1.
KW Signal.
FT CHAIN 1 22 BY SIMILARITY.
FT SIGNAL 23 660 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 660 AA; 70980 MW; 8085BC3CFB46FD3 CRC64;

Query Match 5.6%; Score 102.5; DB 1; Length 660;
Best Local Similarity 19.8%; Pred. No. 0.59; Indels 151; Gaps 20;
Matches 82; Conservative 54; Mismatches 126;
QY 35 TKLL--AIFGLMVLQAG-----ITKPYFVR--AQGLIRACMLYKAGGHVQWA 83
DB 151 TNLVLAAPLSPLPLQDGTNTHIMATEASNAQVRARATIRYRPLVNAVGTAISIS 210
QY 84 FPKLAALGTGYVYHLPLODMAHAGLDLAVAEVPIPSMEVKIITWADTAACGDI 143
DB 211 FWPQTITTSV-----DMNSITSDVRILVPGIASLVI----- 246
QY 144 SGLPVSARRGSEILLGPD--NFEQGWRLLAPI--TAVSOQTRGL-----GCITISLNG 180
DB 247 -----PSERLHYRNQGRSVETSGVAEEETSLVWLCHGSLVNSYTN 290
QY 181 QT-RGLLGI-----ITSLGRDKQ-----VEGEVQVSTAQSEFL 216
DB 291 TPYTGALGLDPALEFPRLTPGNTNTRYRSSTARHRLRGADGTALTTTAATRM 350
QY 217 A-----TCVNGV-----CMTVFH-----GAG-----SKTLAEPKG--PIT 244
DB 351 KDLYFTSTNGVGEIGRGIALTLFNLADTLGLPTELISSAGGLFYSRPVVSANGEPV 410
QY 245 QMTYTNVDQLVGMQAPGARSMTPCTCGSSDLVYV---TRHADVIYVRRRGDSRG--SLIS 300
DB 411 KLYTSVENA-----QDQKGIAPHDIDLGESRVVLDQYDNQHDQRPSPAPSRPFSVL 466
QY 301 PRPVSYLK-----GSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIV 344
DB 467 ANDVLMLSLTAAEYDQSTYGSSTGPFYV--SDSVTLVNVATGAQAVARSLDWTYV 519

DT 01-OCT-1994 (Rel. 30, Last Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Signal recognition particle 54 kDa protein, chloroplast precursor
 DE (SRP54) (54 chloroplast protein) (54CP) (FPC).
 GN FPC OR ATG03940 OR FPF6.150.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RX MEDLINE=94012617; PubMed=8408079;
 RA Franklin A.E., Hoffman N.E.;
 RT "Characterization of a chloroplast homologue of the 54-kDa subunit of
 the signal recognition particle.";
 RL J. Biol. Chem. 268:22175-22180 (1993).
 RP SEQUENCE FROM N.A.
 RP Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;
 RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc,
 encoding the 54 kDa subunit of chloroplast signal recognition
 particle.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneo T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
 RA Nakazaki N., Nambu K., Okumura S., Shino S., Takeuchi C., Wada T.,
 RA Matsumoto M., Yasuda M., Saito S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
 RA Haerum K., Wray J., Johnson D., Rohlfing T., Nelson J.,
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
 RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Voiclaert G., Wambolt R., Duesterhoeft A., Stiekema W., Pohl T.,
 RA Enlian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
 RA Ransperger U., Medler H., Balke K., Wedler E., Peters S.,
 RA van Straalen W., Dirke W., Moolman P., Klein Lankhorst R.,
 RA Weitzenecker T., Botte G., Rose M., Hauf J., Bernerstorfer S., Hempel S.,
 RA Feldpausch W., Lambers S., Villarroel R., Giehl J., Ardiles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:823-826 (2000).
 RL FUNCTION: May target chloroplast proteins to either the thylakoid
 or envelope membranes.
 CC SUBCELLULAR LOCATION: Chloroplast stroma.
 CC TISSUE SPECIFICITY: Most abundant in green shoot tissue and
 CC lower levels seen in the roots and etiolated buds.
 CC SIMILARITY: Belongs to the GTP-binding SRP family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: Z21970; CAA79981.1; --

DR EMBL: AF092168; AAC64139.1; --
 DR EMBL: AL162873; CAB85514.1; --
 DR PIR: S36637; S36637.
 DR HSP: 007347; 1FPH.
 DR InterPro: IPR003593; AAA ATPase.
 DR InterPro: IPR000897; SRP54.
 DR InterPro: IPR004125; SRP54_SBP.
 DR InterPro: IPR004780; SRP_sub.
 DR Pfam: PF00448; SRP54_1.
 DR Pfam: PF02881; SRP54_N; 1.
 DR Pfam: PF02878; SRP_SBP; 1.
 DR ProDom: PD000819; SRP54_1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRFAMs: TIGR00959; ffh; 1.
 DR PROSITE: PS00300; SRP54; 1.
 DR Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;
 KW Transmembrane protein.
 FT TRANSIT 1 75
 FT CHAIN 1 75
 FT DOMAIN 76 370
 FT NP_BIND 371 564
 FT NP_BIND 183 190
 FT NP_BIND 265 269
 FT NP_BIND 323 326
 FT CONFLICT 76 76
 FT SEQUENCE 564 AA; 61232 MW; 423F7285FB963E4 CRC64;
 Query Match 5.5%; Score 101; DB 1; Length 564;
 Best Local Similarity 26.1%; Pred. No. 0.65;
 Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;
 QY 26 VPEI-----DITLLAIIGPMLVLAQI-----TKVPYFAQGLIPACMLVR 72
 DB 154 VDEIVKMGGEVSELQAKSGPVIILAGQGVKTKVACKACVYKQK--KSCMLI- 210
 QY 73 KAAGHYQVAFMKLAAL---TGYVYDHLTPLO--DMAHGLRLAVANPEVIFSDMEV 127
 DB 211 -AGVVFPAIDLVILGSEGVVYTAGTVKADIAKGLKRAK-----NNVDV 261
 QY 128 KIITGADPTACDPIISGLPVSARGREIL-----LGPADFEQGRLLAPITAVSQCT 182
 DB 262 VIM---DIAKRLQIDKMDKDELQVKKFNPTEVLVLDAMTQ--EAAAVTTFNVEI 315
 QY 183 RGLIGCIITSLTGDKNOVEGEVQVS 209
 DB 316 GITGAILTKLDGDSRGGAALSVKVS 341
 RESULT 12
 DPO2_MOUSE
 ID DPO2_MOUSE STANDARD; PRT; 600 AA.
 AC P33611;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
 GN POLA2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1] TAXID=10090;
 RX MEDLINE=93216788; PubMed=8463324;
 RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
 RA Hanaoka F.;
 RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA
 RT polymerase alpha-primase complex and their gene expression during
 RT cell proliferation and the cell cycle.";
 RL J. Biol. Chem. 268:8111-8122 (1993).
 RL FUNCTION: May play an essential role at the early stage of
 CC chromosomal DNA replication by coupling the polymerase

alpha/primase complex to the cellular replication machinery (By similarity).
 CC -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme (subunits A, B, C and D), which is assembled throughout the cell cycle. The largest subunit (subunit A) has DNA polymerase activity, the two smallest subunits (subunits C and D) have DNA primase activity. Subunit B binds to subunit A.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PFM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M PHASE (By Similarity).
 CC -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D13546; BA02746.1; -
 CC PIR: B46642; B46642.
 CC MGD: MGI:99690; P01A2.
 CC InterPro: IPR007200; DNA_pol_alpha_B.
 CC Pfam: PF04058; DNA_pol_alpha_B.1.
 CC DNA replication; Nuclear protein; Phosphorylation.
 CC DOMAIN 101 107 POLY-GU.
 CC FT DOMAIN 115 157 PRO/SER/THR-RICH (CHROMPHILIC).
 CC SEQUENCE 600 AA; 66267 MW; 79F94BEE6F33FEEC CRC64;
 CC -----
 CC Query Match 5.5%; Score 101; DB 1; Length 600;
 CC Best Local Similarity 24.8%; Pred. No. 0.71; Indels 62; Gaps 12;
 CC Matches 55; Conservative 34; Mismatches 71;
 CC -----
 CC 64 LIRACMVRKAAAGHYVM-AFMKALMT-----GTYYVDH-----TPIDMA 106
 CC 27 LAECVIRQEDSMSELNFTSAGKTCITVDILNSFEYVINKLSTAMASASDSC 86
 CC 107 HAGRLDAVAVEPIFSDMEVKIITWGADTPAACDI--ISGLP-----VSARGREI 156
 CC 87 HAGRDI-VSIQELIEAEEETLLSYTPSKPLKRVSTPPTPKRSVAARSPRC- 144
 CC 157 LLGPAADNPEGGMFLAPITVYSCQTKGLGCIITSLTGDKNVBEQVNSTANQSF 216
 CC 145 LLSFSS-----FSPSATPSOK-----YTSRTNR-----GEVVTGSAQ--- 178
 CC 217 ATCVNGVCMVFEHAGSKTL--AGPKGPIITOMTYNVQDILV 256
 CC 179 -----GLSMGSGRGSGSVSLKVPDPELIGSYKAMQQLMG 215
 CC -----
 CC RESULT 13
 CC CYAA LEIDO STANDARD; PRT; 1380 AA.
 CC ID CYAA LEIDO STANDARD; PRT; 1380 AA.
 CC AC Q27675;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Receptor-type adenylylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl cyclase).
 CC GN RAC-A.
 CC OS Leishmania donovani.
 CC OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 CC OX NCBI_TaxID=5661;
 CC RN NCBI_TaxID=5661;
 CC RP SEQUENCE FROM N.A.A.
 CC RC STRAIN=1S Sudanese;
 CC RX MEDLINE=95340554; PubMed=7615561;
 CC RA Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.,
 CC RT "A family of putative receptor-adenylylate cyclases from Leishmania donovani".
 CC RL J. Biol. Chem. 270:17551-17558(1995).
 CC -1- FUNCTION: Could act as a receptor for a unknown ligand.

CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
 CC -1- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.
 CC -1- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote) but not in the mammalian host stage of the parasite life cycle.
 CC -1- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U17042; AAA74998.1; -
 CC PIR: T18309; T18309.
 CC InterPro: IPR001054; G_cyclase.
 CC Pfam: PF00211; guanylate_cyc; 1.
 CC SMART: SM00044; CYCC; 1.
 CC DR PROSITE: PS50125; GUANYLATE CYCLASES_2; 1.
 CC DR Lyase; CAMP biosynthesis; Transmembrane; Receptor; Glycoprotein;
 CC Metal-binding; Magnesium.
 CC DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 35 55 POTENTIAL.
 CC DOMAIN 56 891 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 892 912 POTENTIAL.
 CC DOMAIN 913 1380 CYTOPLASMIC (POTENTIAL).
 CC METAL 938 938 MAGNESIUM (BY SIMILARITY).
 CC METAL 981 981 MAGNESIUM (BY SIMILARITY).
 CC CARBOHYD 422 422 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 478 478 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 497 497 N-LINKED (GLCNAC...) (POTENTIAL).
 CC CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL).
 CC SEQUENCE 1380 AA; 151692 MW; 6B2DF7D3C1107A0 CRC64;
 CC -----
 CC Query Match 5.1%; Score 93.5; DB 1; Length 1380;
 CC Best Local Similarity 21.0%; Pred. No. 8.4; Indels 159; Gaps 24;
 CC Matches 93; Conservative 51; Mismatches 139;
 CC -----
 CC 14 GGRDALITLTCAVHP---ELIFDITKL-----LIALFGPLM-----V 47
 CC 101 GGPPIKITL---HPDPQDNLYDAEVLHSLAQEKILAVLGYLDRLTALPSNADY 155
 CC 48 LQAGITKVP-----YFVRQGLIRACMVIRKAAAGHYV-----QWAFMK 86
 CC 156 YOSGLMLLAPFTGSSGVRTWSVYFTRAEPMVELKVLVLM-----HIVRLARRVAFMR 210
 CC 87 LIALTG-----TYVDLTLPLQMAHAGRLDAVA--VEPIFSDMEV----- 128
 CC 211 ---LTGMHFGGELTYVQDTLSL-----LRDPALVLTVPYSSESVVEDEAFDAVAD 260
 CC 129 -----ITWGADTPAACGDIISGLPVSARGREILG-----PANDFEG 167
 CC 261 TNPQVITWAAPVQVYIFLEKLTDPRTSSAVYVSCSMIQRVEDVYRLLSAGIKPO 320
 CC 168 GWRLLAPITAVSQTRGLGCIITSLTGDRKNQVEG-----VQVNSTATOSFLATCVNGV 223
 CC 321 DGRILASATL-----SPVSEGLKWEVLKAAQMSNVIE---NSG 356
 CC 224 CWTVEHAGSKTLAGEK---GPIITMYNVDDLVGMQAPPGARSMTECTGSSLLIYV 279
 CC 357 SPDYPPDDSTETLIGRKARSEAPLSRKYT-VDE---FQAHPSLIAKMLGMLSGTLVQO 412
 CC 280 TRHADVIVPARRDSDGSLSPRPV---SYLKGSSGGLIC-PSGAVGIFPRAVCTRGV 335
 CC 413 TLEQTMVYKRSYKAGLGNQRNFVIGSDVVLGDIYGP--CEFLAFLG--ASCYCNQSG 468
 CC 336 AKAV-----DFIPVESME 348
 CC 469 HSSILITVLQNASWDIVPDSSEK 490

RESULT 14
 ID TOLB_CHLITE STANDARD; PRT; 434 AA.
 AC Q8K600;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE TOLB protein precursor.
 GN TOLB OR C70636.
 OS Chlorobium tepidum.
 CC Bacteria; Chlorobi; Chlorobiales; Chlorobiaceae;
 CC Chlorobium tepidum.
 CC Chlorobium.
 CC NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TL5 / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vanatvean J., Khouli H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
 RT "The complete genome sequence of Chlorobium tepidum TLS, a
 photocynthetic, anaerobic, green-sulfur bacterium.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
 CC -1- FUNCTION: Involved in the tomb-independent uptake of proteins (By
 similarity).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -1- SIMILARITY: Belongs to the tolb family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AEO12837; AAM71875.1; -.
 DR TIGR; CT0636; -.
 DR HAMAP; MF 00671; -; 1.
 KM Transport; Protein transport; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 27 Potential.
 FT CHAIN 28 434 TOLB protein.
 SQ SEQUENCE 434 AA; 47275 MW; 71A347B89C7A0F99 CRC64;
 Query Match 5.0%; Score 93; DB 1; Length 434;
 Best Local Similarity 21.2%; Pred. No. 2.3;
 Matches 69; Conservative 47; Mismatches 139; Indels 70; Gaps 14;
 Db 42 FGIIMVLOAGITVPPFVRAGLIRACMLVRKAGHYQAMFKLAALTGYVDHLTP 101
 8 FACLCIMLFQMLFVPTLRAB-VGEYIAR-K-EGASRIAVLIDKTSADGGKQ----- 58
 Db 102 LQDMAA-----GLRLAAVEVEVPSDMVKIITGADTAACGDIISGLPVS 149
 59 -REMASSLDVTINKGIDFTGLFRLRA--PLNIRNQNGNINPASIAGGIIYAGGSYT 115
 Db 150 ARRGREIL-----LQPADNFEQGMRLIAPITVASSQOTRGLGCTITS 192
 116 KRSGREPLEMHWYSSGSKSLARTYTGESQLEAIGLRCADIVELLTKRESVGRIVAF 175
 Db 193 LTRGRDKNQ-----VEGEVQVSTATOSFLATCV-----NGVCWVYEHGASGKTLAAGPG 241
 176 VANRTGNKEIYMDPFGENVVQLTNSRSISLTPAVSPDGTIYAMDY--TSKPNLIYKN 233
 Db 242 PITOMTNNVDQDLY---GWAQAPGARSMTPCTC--GSSDLYLVTRHADVIFVRERGDSP 295
 234 IATGAKVAVKRGVCISPAWR--PGTNTLVTTLSYEGDQDLYLI--RADGTVERRRLTKGG 289
 Db 296 GSILSP-----RPVSLKSGSGCP 314

Db 290 GIDVSPFSPDGSXMAFVSTRQSGP 314
 RESULT 15
 ID TRFE_HORSE STANDARD; PRT; 706 AA.
 AC P27425;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sero transferrin precursor (transferrin) (Siderophilin) (Beta-1-metal
 binding globulin).
 GN TF.
 OS Equus caballus (Horse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 CC NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=93277958; PubMed=8504171;
 RA Carpenter M.A., Broad T.E.;
 RT "The cDNA sequence of horse transferrin.";
 RL Biochim. Biophys. Acta 1173:230-232(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Extraembryonic tissue;
 RC McDowell K.J., Adams M.H., Baker C.B.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Transferrins are iron binding transport proteins which
 CC can bind two atoms of ferric iron in association with the binding
 CC of an anion, usually bicarbonate. It is responsible for the
 CC transport of iron from sites of absorption and heme degradation to
 CC those of storage and utilization. Serum transferrin may also have
 CC a further role in stimulating cell proliferation.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
 CC -1- DOMAIN: Composed of two homologous domains.
 CC -1- SIMILARITY: Belongs to the transferrin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M69020; AAA30958.1; -.
 DR EMBL; U21127; AAA63684.1; -.
 DR PIR; S33761; S33761.
 DR HSRF; P02787; IABE.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PRO0422; TRANSFERRIN.
 DR SMART; SMO0094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KM Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 706 SEROTRANSFERRIN.
 FT REPEAT 20 357 1.
 FT REPEAT 358 706 2.
 FT DISULFD 26 64 BY SIMILARITY.
 FT DISULFD 36 55 BY SIMILARITY.
 FT DISULFD 134 215 BY SIMILARITY.
 FT DISULFD 174 190 BY SIMILARITY.
 FT DISULFD 177 198 BY SIMILARITY.
 FT DISULFD 187 200 BY SIMILARITY.
 FT DISULFD 248 262 BY SIMILARITY.

```
FT DISULFID 360 623 BY SIMILARITY.
FT DISULFID 366 398 BY SIMILARITY.
FT DISULFID 376 389 BY SIMILARITY.
FT DISULFID 423 701 BY SIMILARITY.
FT DISULFID 441 664 BY SIMILARITY.
FT DISULFID 474 550 BY SIMILARITY.
FT DISULFID 498 692 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.
FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
FT BINDING 143 143 SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT BINDING 483 483 SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;
```

Query Match 5.0%; Score 92.5; DB 1; Length 706;
Best Local Similarity 21.5%; Pred. No. 4.5;
Matches 64; Conservative 43; Mismatches 109; Indels 81; Gaps 17;

```
QY 82 MAFKALALGTYY---YDHLTPLODMAHGLRDLAVAVEPIFSDMEVKITWGA----134
Db 321 LGFLRIPFPAADTWLYLGEYVT-----AIRVLREDIRPEVKD-ECKVVMCAIGHH 371
QY 135 DTAACGD-IISGLPVASARRGR-----ELLGPAUNFEGQGWRL-----LAPITAY 178
Db 372 EKVCDEMVSNGNICECSAQSTEDCIKIVKGEADAMSLDGGFIYIAGKCGIPLVLAH 431
QY 179 SQCTRGILGCIITSLTRDKQVGEVQVSTATQSFLATCVNGVCTVPHGASKTLAG 238
Db 432 NYETRSSACVDTPPEGTH-----AAVAVKSSSDPDLT-----NSLKG 470
QY 239 PKGPIITOMYITVDDPLVGMQAPPGARSMTPCTCGSSDLYLVTRADVIPIVRRGDSRSL 298
Db 471 KK-----SCHTCVDR-TAGWNIPMGL-----LYSEIKHCEFDKFFREGCAGYR 513
QY 299 LSPRPVITLKSSSGGP-LLC-PSGHA-----VGIFRAVCTRGVAKAVDFIPVESME 348
Db 514 RNSLTCLNLCIGSAGSPGSECEPNNHRYGYGTGAFCLVEKGDVA---FVKQOTVE 566
```

Search completed: May 6, 2004, 09:31:53
Job time : 8.34864 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 32.9219 Seconds
(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-13
Perfect score: 1842
Sequence: 1 AHQVATPPPLNFGGRDAII.....RGVAKAVDPFVPSMETTKR 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phase:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaeoprotein:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	97.3	3010	12 Q9DTE6	Q9dte6 hepatitis c
2	1784	96.9	3010	12 Q9J3F9	Q9j3f9 hepatitis c
3	1783	96.8	3010	12 P90191	P90191 hepatitis c
4	1782	96.7	3010	12 Q9J3H7	Q9j3h7 hepatitis c
5	1781	96.7	3010	12 Q9J3H5	Q9j3h5 hepatitis c
6	1780	96.6	3010	12 Q9DTE6	Q9dte6 hepatitis c
7	1778	96.5	3010	12 Q9J3F4	Q9j3f4 hepatitis c
8	1778	96.5	3010	12 P88803	P88803 hepatitis c
9	1775	96.4	3010	12 Q68788	Q68788 hepatitis c
10	1775	96.4	3010	12 Q807P3	Q807P3 hepatitis c
11	1774	96.3	3010	12 Q81755	Q81755 hepatitis c
12	1774	96.3	3010	12 Q81817	Q81817 hepatitis c
13	1774	96.3	3010	12 P89366	P89366 hepatitis c
14	1774	96.3	3010	12 Q9J3G6	Q9j3g6 hepatitis c
15	1773	96.3	3010	12 Q99AU2	Q99au2 hepatitis c
16	1773	96.3	3010	12 Q9J3H3	Q9j3h3 hepatitis c

17	1773	96.3	3010	12 Q9QIX6	Q9qix6 hepatitis c
18	1773	96.3	3010	12 Q9QIX5	Q9qix5 hepatitis c
19	1773	96.3	3010	12 Q9DTE4	Q9dte4 hepatitis c
20	1772	96.2	3010	12 Q9DTE7	Q9dte7 hepatitis c
21	1771	96.1	3010	12 Q9J3I0	Q9j3i0 hepatitis c
22	1770	96.1	3010	12 Q9QIX3	Q9qix3 hepatitis c
23	1770	96.1	3010	12 Q09796	Q09796 hepatitis c
24	1769	96.0	3010	12 Q9QIX5	Q9qix5 hepatitis c
25	1769	96.0	3014	12 Q9DTE0	Q9dte0 hepatitis c
26	1768	96.0	3010	12 Q9DTE0	Q9dte0 hepatitis c
27	1767	95.9	3010	12 Q9J3H9	Q9j3h9 hepatitis c
28	1766	95.9	3010	12 Q9J3H0	Q9j3h0 hepatitis c
29	1766	95.9	3010	12 Q68826	Q68826 hepatitis c
30	1766	95.9	3010	12 Q9WMX2	Q9wmx2 hepatitis c
31	1765	95.8	3010	12 Q81760	Q81760 hepatitis c
32	1765	95.8	3010	12 Q9QIX8	Q9qix8 hepatitis c
33	1765	95.8	3010	12 Q9QIX7	Q9qix7 hepatitis c
34	1764	95.8	3010	12 Q02829	Q02829 hepatitis c
35	1764	95.8	3010	12 Q9J3H6	Q9j3h6 hepatitis c
36	1763	95.7	3010	12 Q9J3I1	Q9j3i1 hepatitis c
37	1761	95.6	3010	12 Q9QIT4	Q9qit4 hepatitis c
38	1761	95.6	3010	12 Q9DTE3	Q9dte3 hepatitis c
39	1761	95.6	3011	12 Q9J3H4	Q9j3h4 hepatitis c
40	1759	95.5	3013	12 Q9J3I6	Q9j3i6 hepatitis c
41	1758	95.4	3010	12 Q8QRL8	Q8qrl8 hepatitis c
42	1757	95.4	3010	12 Q81541	Q81541 hepatitis c
43	1757	95.4	3010	12 Q92973	Q92973 hepatitis c
44	1755	95.3	2864	12 Q9DTE5	Q9dte5 hepatitis c
45	1755	95.3	3010	12 Q9DTE5	Q9dte5 hepatitis c

ALIGNMENTS

RESULT 1
ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.
AC Q9DTE6;
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage, Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV142;
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
with hepatocellular carcinoma: the 'progression score' revisited.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL: AB049091; BAB1804.1; -
CC PIR: A61196; A61196.
CC PIR: P50329; P50329.
CC HSSP: P26663; IUXP.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR GO: GO:0019028; C:viral capsid; IEA.
DR GO: GO:0019031; C:viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F:electon transporter activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003688; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006118; P:electron transport; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR003003; Cys_ser_lysln.
 DR InterPro: IPR000345; Cys_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVlr.
 DR Pfam: PF01543; HCV_core; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01538; HCV_NSI; 1.
 DR Pfam: PF02907; HCV_NS2; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF000998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydroxase; Nonstructural protein; Polymerase; Transmembrane.
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SC SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 97.3%; Score 1792; DB 12; Length 3010;
 Best Local Similarity 95.7%; Pred. No. 6,3e-143;
 Matches 337; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

DR 1 AHLOWPPLNNGGRRAIILITCAVHPELIPITIKLLAIGPILVLOAGITKVPYFVR 60
 DB 855 AHLOWPPLNNGGRRAIILITCAVHPELIPITIKLLAIGPILVLOAGITRVPYFVR 914
 QY 61 AAGLIRACMLVRAAGSHYQMAFMKLAALTGTYYVDHLTPLODMAHAGIRDLAIVAEV 120
 DB 915 AAGLIRACMLVRAAGSHYQMAFMKLAALTGTYYVDHLTPLODMAHAGIRDLAIVAEV 974
 QY 121 IFSDMEVKITTWGADTPACDIIISGLPVARRGREILGPDNFBEGQWRLAFTIAYQ 180
 DB 975 VFSDMEKITTWGADTPACDIIISGLPVARRGREILGPDNFBEGQWRLAFTIAYQ 1034
 QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFIATCNGVCMVFVFGAGSKTLAIGR 240
 DB 1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFIATCNGVCMVFVFGAGSKTLAIGR 1094
 QY 241 GPITQWNTNVDOLVGMQAPPGARSMTCTCGSSDIYLVTRADVIIVRRGDSRGLIS 300
 DB 1095 GPITQWNTNVDOLVGMQAPPGARSMTCTCGSSDIYLVTRADVIIVRRGDSRGLIS 1154
 QY 301 PRVSVYIKGSSGGLPLCPGSHAVIGFRAACTCGVAKAVDFIVEMETMR 352
 DB 1155 PRVSVYIKGSSGGLPLCPGSHAVIGFRAACTCGVAKAVDFIVEMETMR 1206

RESULT 2

Q933F9 ID Q933F9 PRELIMINARY; PRT; 3010 AA.
 AC Q933F9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCB TaxID=11103;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=MD3.
 CC Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;
 CC "Characteristics of hepatitis C viral genome associated with disease
 CC progression.";
 CC Submitted (Nov-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: '
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; AF207774; AAF65964.1; -
 CC PIR; A61196; A61196.
 CC PIR; P0246; P0246.
 CC PIR; PS0329; PS0329.
 CC HSR; P27958; 1HE1.
 CC MEROPS; S29.001; -.
 DR MEROPS; U39.001; -.
 DR GO:0016021; C:integral to membrane; IEA.
 DR GO:0019028; C:viral capsid; IEA.
 DR GO:0019031; C:viral envelope; IEA.
 DR GO:0005524; F:ATP binding; IEA.
 DR GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO:0005489; F:electron transporter activity; IEA.
 DR GO:0003723; F:RNA binding; IEA.
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006118; F:electron transport; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser trypsin.
 DR InterPro: IPR00345; Cys_ser_lysln.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVlr.
 DR Pfam: PF01543; HCV_core; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01538; HCV_NSI; 1.
 DR Pfam: PF02907; HCV_NS2; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF000998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NSI; 1.

DR SMART, SM00487, DEXDC, 1.
 DR PROSITE, PS00190, CYTOCHROME C, 1.
 KM Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327102 MW; 7162CDB93E60C7 CRC64;

Query Match 96.8%; Score 1784; DB 12; Length 3010;
 Best Local Similarity 95.2%; Pred. No. 3e-142;
 Matches 335; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAIILITCAVPELIPDITKLLAIFGPIWVQAGITKVPYFVR 60
 DB 855 AHLQWIPPLNVRGGRDAIILITCAVPELIPDITKLLAIFGPIWVQAGITKVPYFVR 914

QY 61 AAGLIRACMLVRKAGGHHYQMAFMKALITGYVYDHTLPLODMAHGLRLAIAVEPV 120
 DB 915 AAGLIRACMLVRKAGGHHYQMAFMKALITGYVYDHTLPLODMAHGLRLAIAVEPV 974

QY 121 IFSMEVKIITWGADTAACGDIISGLPVASRGREILGADNFEQGGRLLAPITAYSQ 180
 DB 975 VFSDEMKIITWGADTAACGDIISGLPVASRGREILGADNFEQGGRLLAPITAYSQ 1034

QY 181 QTRGLGCIITSLTGRDKNQEVEGVVSTAFOSFLATCVNGVCMVFFHAGSKTLAIGPK 240
 DB 1035 QTRGLGCIITSLTGRDKNQEVEGVVSTAFOSFLATCVNGVCMVFFHAGSKTLAIGPK 1094

QY 241 GPITQMTYNDQDVLVGMQAPPGARSMPTCTGSSDLVLTREHADVTPVRRRGRSGSLLS 300
 DB 1095 GPITQMTYNDQDVLVGMQAPPGARSMPTCTGSSDLVLTREHADVTPVRRRGRSGSLLS 1154

QY 301 PRPVSYKGGSSGGLPCPSGAVGIFRAVCTRGVAKAVPIPVESMETMR 352
 DB 1155 PRPVSYKGGSSGGLPCPSGAVGIFRAVCTRGVAKAVPIPVESMETMR 1206

RESULT 3
 P90191 PRELIMINARY; PRT; 3010 AA.

ID P90191
 AC P90191
 DT 01-MAY-1997 (TRENBLREL. 03, Created)
 DT 01-MAY-1997 (TRENBLREL. 03, Last sequence update)
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV-1b;
 RA Enomoto N.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV-1b;
 RA MEDLINE=95340824; PubMed=7542279;
 RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
 RA Yamamoto C., Izumi N., Marumo F., Sato C.;
 RT "Comparison of full-length sequences of interferon-sensitive and
 RT resistant hepatitis C virus 1b";
 RL J. Clin. Invest. 96:224-230 (1995).
 CC -1- SUBUNIT: THE VARIATION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL: D50482; BAA09073.1; -.
 DR PIR: A61196; A61196.
 DR PIR: P00254; P00254.
 DR PIR: P00804; P00804.
 DR PIR: P50329; P50329.
 DR PDB: 1DXW; 12-JAN-01.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.

DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0001968; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_cysteine.
 DR InterPro: IPR001410; DAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_core.
 DR InterPro: IPR002511; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001450; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_D5_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01538; HCV_NS1; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART, SM00487, DEXDC, 1.
 KM Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1972
 FT CHAIN 1973 2419
 FT CHAIN 2420 3010
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match 96.8%; Score 1783; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 3.7e-142;
 Matches 336; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAIILITCAVPELIPDITKLLAIFGPIWVQAGITKVPYFVR 60
 DB 855 AHLQWIPPLNVRGGRDAIILITCAVPELIPDITKLLAIFGPIWVQAGITKVPYFVR 914

QY 61 AAGLIRACMLVRKAGGHHYQMAFMKALITGYVYDHTLPLODMAHGLRLAIAVEPV 120
 DB 915 AAGLIRACMLVRKAGGHHYQMAFMKALITGYVYDHTLPLODMAHGLRLAIAVEPV 974

QY 121 IFSMEVKIITWGADTAACGDIISGLPVASRGREILGADNFEQGGRLLAPITAYSQ 180
 DB 975 VFSDEMKIITWGADTAACGDIISGLPVASRGREILGADNFEQGGRLLAPITAYSQ 1034

QY 181 QTRGLGCIITSLTGRDKNQEVEGVVSTAFOSFLATCVNGVCMVFFHAGSKTLAIGPK 240
 DB 1035 QTRGLGCIITSLTGRDKNQEVEGVVSTAFOSFLATCVNGVCMVFFHAGSKTLAIGPK 1094

QY 241 GPTQMTNVDQIVGQAPPGASMTPTCTGSSDLVLTNRADVTPVRRGDSRGLS 300
 DB 1095 GPTQMTNVDQIVGQAPPGASMTPTCTGSSDLVLTNRADVTPVRRGDSRGLS 1154
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 352
 DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 1206

RESULT 4

Q9J3H7 PRELIMINARY; PRT; 3010 AA.
 AC Q9J3H7;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBT_Taxid=11103;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=MD15;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 progression."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL: AF207756; AAF65946.1; -
 DR PIR: A61196; A61196.
 DR PIR: P00246; P00246.
 DR PIR: P00804; P00804.
 DR PIR: P50329; P50329.
 DR HSP: P26663; IXP.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006118; F:electron transport; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P:transcription; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_lysine.
 DR InterPro: IPR000345; CytC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_NS1.
 DR InterPro: IPR002531; HCV_NS2.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR00745; HCV_NS4.
 DR InterPro: IPR001490; HCV_NS4.
 DR InterPro: IPR002868; HCV_NS5.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR004109; peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PStir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.

DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR ProDom: PD086062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR ProSITE: PS00190; CYTOCHROME_C; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327365 MW; D8653F7317FFA106 CRC64;

Query Match

Best Local Similarity 96.7%; Score 1782; DB 12; Length 3010;
 Matches 334; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHLQWVPLNVRGGRDAIILLTCVAPPELLFDITKLLAIFGLMTLQAGITKVPYFVR 60
 DB 855 AHLQWVPLNVRGGRDAIILLTCVAPPELLFDITKLLAIFGLMTLQAGITKVPYFVR 914
 QY 61 AGLIRACMLVRKKAAGHYVQMAFMKLAALTGTVYVHLPLOMAGLRLAVAEV 120
 DB 915 AGLIRACMLVRKKAAGHYVQMAFMKLAALTGTVYVHLPLOMAGLRLAVAEV 974
 QY 121 IFSDEVEKIIITWGAADTAACDIIISGLPVSARGREIILGPADNFEQGWRLAPITAYSQ 180
 DB 975 IFSDEVEKIIITWGAADTAACDIIISGLPVSARGREIILGPADNFEQGWRLAPITAYSQ 1034
 QY 161 QTRGLGLGIIITSLGRDNQYEGEVYSTQSFPLATCVNGVWTYFHGAGSTLTAQPK 240
 DB 1035 QTRGLGLGIIITSLGRDNQYEGEVYSTQSFPLATCVNGVWTYFHGAGSTLTAQPK 1094
 QY 241 GPTQMTNVDQIVGQAPPGASMTPTCTGSSDLVLTNRADVTPVRRGDSRGLS 300
 DB 1095 GPTQMTNVDQIVGQAPPGASMTPTCTGSSDLVLTNRADVTPVRRGDSRGLS 1154
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 352
 DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 1206

RESULT 5

Q9J3H5 PRELIMINARY; PRT; 3010 AA.
 AC Q9J3H5;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBT_Taxid=11103;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=MD17;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 progression."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL: AF207758; AAF65948.1; -
 DR PIR: A61196; A61196.
 DR PIR: P00246; P00246.
 DR PIR: P00804; P00804.
 DR PIR: P50329; P50329.
 DR HSP: P27958; IHEI.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.

DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0016787; F:hydrolyase activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P:transcription; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR003003; Cys Ser tyrosin.
 DR InterPro: IPR000345; CytC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR Pfam: PF0186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KM Hydrolyase; Nonstructural protein; Polyprotein;
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326801 MW; 9FEE3D1B93B7AA4B CRC64;

Query Match 96.7%; Score 1781; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 5.5e-142;
 Matches 336; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
 QY 1 AHQVMPINPVRNRCGRDAIILLTCVAPREIFDITKLLAIFGLMVLQGITKVPYFVR 60
 DB 855 AHQVMPINPVRNRCGRDAIILLTCVAPREIFDITKLLAIFGLMVLQGITRVPYFVR 914
 QY 61 AAGIIRACMLVRKAGHYOMAFMKALATGTYVYHPLDMDMAGLRDAVAPEV 120
 DB 915 AAGIIRACMLVRKAGHYOMAFMKALATGTYVYHPLDMDMAGLRDAVAPEV 974
 QY 121 IFSDMVEKILTWGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 180
 DB 975 VFSMEKILTWGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 1034
 QY 181 QTRRLGCIITSLTGKDNQVVEGVVYATQSFATCNGVCWYTFHAGSKTLAEPK 240
 DB 1035 QTRRLGCIITSLTGKDNQVVEGVVYATQSFATCNGVCWYTFHAGSKTLAEPK 1094
 QY 241 GPIITQMTNVDDLVGVQAPPGASMTPTCGSSDLYLVTRHADVIFVRRRGDSRGLLS 300
 DB 1095 GPIITQMTNVDDLVGVQAPPGASMTPTCGSSDLYLVTRHADVIFVRRRGDSRGLLS 1154

QY 301 PEPVSYLKSSGGPLCPGSHAVGIFRAVCTRGVAKAVIDEIPVESMETTR 352
 DB 1155 PEPVSYLKSSGGPLCPGSHAVGIFRAVCTRGVAKAVIDEIPVESMETTR 1206
 RESULT 6
 ID Q9DTE6 PRELIMINARY; PRT; 3010 AA.
 AC Q9DTE6
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA, positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV221;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
 RA Mishihiro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 RT with hepatocellular carcinoma: the 'progression score' revisited";
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND NS5A (BY SIMILARITY).
 DR EMBL: AB049101; BAB18814.1; --
 DR PIR: A61196; A61196.
 DR PIR: PQ0246; PQ0246.
 DR PIR: PS0329; PS0329.
 DR HSSP: P26653; IJXP.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019026; C:viral capsid; IEA.
 DR GO: GO:0018031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0016787; F:hydrolyase activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR GO: GO:0006350; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006508; P:transcription; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_tyrosin.
 DR InterPro: IPR000345; CytC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.

DR Pfam; PF01538; HCV NS3; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR Pfam; PF01602; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KM Hydrolyase; Nonstructural protein; Polypeptide;
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SO SEQUENCE 3010 AA; 327108 MW; DE182D810EF79EE4 CRC64;

Query Match 96.5%; Score 1780; DB 12; Length 3010;
 Best Local Similarity 95.7%; Pred. No. 6.6e-142;
 Matches 337; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 AHQWVPLNVRGGRDAIILLTCVAPHELIIFDITKLLAIFGRLVWLQNGITKVEFVR 60
 DB 855 AHLHWVPLNVRGGRDAIILLTCVAPHELIIFDITKLLAIFGRLVWLQNGITKVEFVR 914
 QY 61 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLLPLQDMAHAGRLDAVAEVPV 120
 DB 915 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLLPLQDMAHAGRLDAVAEVPV 974
 QY 121 IFSMEVETITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 180
 DB 975 VFSMEVETITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 1034
 QY 181 QTRGLGCIITSLTGRDNKVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 240
 DB 1035 QTRGLGCIITSLTGRDNKVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 1094
 QY 241 GPTQMTNTNDOLVGMQAPRGASMTPTCCGSSDLYLVRHADVIFVRRGDSRGLS 300
 DB 1095 GPTQMTNTNDOLVGMQAPRGASMTPTCCGSSDLYLVRHADVIFVRRGDSRGLS 1154
 QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIVESMETTMR 352
 DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 7
 Q9J3F4 PRELIMINARY; PRT; 3008 AA.
 ID Q9J3F4
 AC Q9J3F4
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OS Hepatitis C virus.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis C virus.
 OC NCBI_Taxid=1103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD34;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 progression";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBMITT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC EMBL; AF208024; AAF61205.1; -;
 DR PIR; A61196; A61196;
 DR PIR; P00246; P00246;
 DR PIR; P03229; P03229;
 DR HSP; P26653; IUXP.

DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F: electron transporter activity; IEA.
 DR GO; GO:0016787; F: hydrolyase activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008235; F: serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006118; F: electron transport; IEA.
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P: translation; IEA.
 DR GO; GO:0019079; P: viral genome replication; IEA.
 DR GO; GO:0019087; P: viral transformation; IEA.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR000345; Cys_heme_BS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002688; HCV NS4b.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR004109; peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF01537; HCV NS3; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00998; Viral RdRp; 1.
 DR Pfam; PF01602; HCV NS1; 1.
 DR PRODOM; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 KM ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KM Hydrolyase; Nonstructural protein; Polypeptide;
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SO SEQUENCE 3008 AA; 326834 MW; 99AB09E14C3109F4 CRC64;

Query Match 96.5%; Score 1778; DB 12; Length 3008;
 Best Local Similarity 95.5%; Pred. No. 9.8e-142;
 Matches 336; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHQWVPLNVRGGRDAIILLTCVAPHELIIFDITKLLAIFGRLVWLQNGITKVEFVR 60
 DB 853 AHLHWVPLNVRGGRDAIILLTCVAPHELIIFDITKLLAIFGRLVWLQNGITKVEFVR 912
 QY 61 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLLPLQDMAHAGRLDAVAEVPV 120
 DB 913 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLLPLQDMAHAGRLDAVAEVPV 972
 QY 121 IFSMEVETITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 180
 DB 973 VFSMEVETITWGDITACGDIISGLPVSARREIILGPADNFEQGMRLAPITAYSQ 1032
 QY 181 QTRGLGCIITSLTGRDNKVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 240
 DB 1033 QTRGLGCIITSLTGRDNKVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAGPK 1032
 QY 241 GPTQMTNTNDOLVGMQAPRGASMTPTCCGSSDLYLVRHADVIFVRRGDSRGLS 300

Db 1093 GPTTQMTNVDDLVGMQAPPGARSLTPTCGSSDLYLTRADYIPVRRRDSRGSLLS 1152
 QY 301 PRPVSYLKSSGGGFLCPGSHAVGIFRAVCTRGVAKAVDFPVESMETTMR 352
 Db 1153 PRPVSYLKSSGGGFLCPGSHAVGIFRAVCTRGVAKAVDFPVESMETTMR 1204

RESULT 8
 P88803 PRELIMINARY; PRT; 3010 AA.
 ID P88803
 AC P88803;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV-1b;
 RA Enomoto N.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV-1b;
 RA MEDLINE=95340824; PubMed=7542279;
 RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
 RA Yamamoto C., Izumi N., Marumo F., Sato C.;
 RT "Comparison of full-length sequences of interferon-sensitive and
 RT resistant hepatitis C virus 1b.";
 RL J. Clin. Invest. 96:224-230 (1995).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL: D50484; BAA09075.1; -
 DR PIR: A61196; A61196.
 DR HSP: P25663; INS3.
 DR GO: GO:0016021; C: integral to membrane; IEA.
 DR GO: GO:0019028; C: viral capsid; IEA.
 DR GO: GO:0019031; C: viral envelope; IEA.
 DR GO: GO:0005524; F: ATP binding; IEA.
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO: GO:0003723; F: RNA binding; IEA.
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F: structural molecule activity; IEA.
 DR GO: GO:0016740; F: transferase activity; IEA.
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P: transcription; IEA.
 DR GO: GO:0019079; P: viral genome replication; IEA.
 DR GO: GO:0019087; P: viral transformation; IEA.
 DR InterPro: IPR009003; Cys_ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5b.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.

DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02807; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF018602; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polypeptide; RNA-directed RNA polymerase; transferase; Transmembrane.
 FT CHAIN 1
 FT CHAIN 191
 FT CHAIN 192 383
 FT CHAIN 384 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1972
 FT CHAIN 1973 2419
 FT CHAIN 2420 3010
 SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783FEFB8 CRC64;
 Query Match 96.5%; Score 1778; DB 12; Length 3010;
 Best Local Similarity 94.9%; Pred. No. 9, 8e-142;
 Matches 334; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 AHQWIPPLNVRGGRDALITLCAVHPELIPDITKLLAIFGPMVLAQGITKYFVR 60
 Db 855 AHQWIPPLNVRGGRDALITLCAVHPELIPDITKLLAIFGPMVLAQGITKYFVR 914
 QY 61 AAGLIRACMLVKAAGHYQMAFKLAALGTYYDHLTPLODMAHAGRLDAVAEVR 120
 Db 915 AAGLIRACMLVKAAGHYQMAFKLAALGTYYDHLTPLODMAHAGRLDAVAEVR 974
 QY 121 ISDMVKIITGAPLTAAGDIIISGLPVARGREILIGPANEFGQGRLLAPITAYSO 180
 Db 975 VSDMETKIITGADPLAACGDIISGLPVARGREILIGPANEFGQGRLLAPITAYSO 1034
 QY 181 QTRGLGCIITLTGDKNQVEGEVQVSTATQSFPLATCVNGVCTVFHAGSKTLAPK 240
 Db 1035 QTRGLGCIITLTGDKNQVEGEVQVSTATQSFPLATCVNGVCTVFHAGSKTLAPK 1094
 QY 241 GPTTQMTNVDDLVGMQAPPGARSLTPTCGSSDLYLTRADYIPVRRRDSRGSLLS 300
 Db 1095 GPTTQMTNVDDLVGMQAPPGARSLTPTCGSSDLYLTRADYIPVRRRDSRGSLLS 1154
 QY 301 PRPVSYLKSSGGGFLCPGSHAVGIFRAVCTRGVAKAVDFPVESMETTMR 352
 Db 1155 PRPVSYLKSSGGGFLCPGSHAVGIFRAVCTRGVAKAVDFPVESMETTMR 1206

RESULT 9
 O68788 PRELIMINARY; PRT; 3010 AA.
 ID O68788
 AC O68788;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE HCV polypeptide (Genome polypeptide).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96362158; PubMed=8720135;
 RA Seki M., Honda Y.;
 RT "Phosphotransferase antisense oligodeoxynucleotides capable of
 RT inhibiting Hepatitis C virus gene expression: In vitro translation
 RT assay.";
 RL J. Biochem. 118:1199-1204 (1995).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

Qy	24	GPTQMTVTDVDDVGNQMAPPCARSMETCTGSSDVLVLTVRHADVIPVRGRGSRSLLS	300
Db	1095	GPTQMTVTDVDDVGNQMAPPCARSMETCTGSSDVLVLTVRHADVIPVRGRGSRSLLS	115
Qy	301	PRPVSYLKSSGGPPLICPSGNAVGI.FRAAVTCRGVAAVDFIPVESMETTW	352
Db	1155	PRISYLYKSSGGPPLICPSGHVVGIFRAAVTCRGVAAVDFIPVESMETTW	1206
RESULT 10.			
Q807P3	PRELIMINARY	FRT, 3010 AA.	
AC	Q807P3		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Polypeptide.		
OS	Hepatitis C virus.		
OC	Vitnutes; ssRNA positive-strand vitnutes, no DNA seage, Flaviviridae;		
OC	Hepacivirus.		
OX	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MILE;		
RX	MEDLINE=22047193; PubMed=12051758;		
RA	Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,		
RA	Nio Y., Hosaka M., Miyahara Y., Shimotohno K.;		
RT	"Subgenomic replicon derived from a cell line infected with the		
RT	hepatitis C virus.";		
RL	Biochem. Biophys. Res. Commun. 293:993-999(2002).		
DR	EMBL; AB080299; BACS4896.1; -		
DR	GO: GO:0019029; C:viral capsid; IEA.		
DR	GO: GO:0019031; C:viral envelope; IEA.		
DR	GO: GO:0005524; F:ATP binding; IEA.		
DR	GO: GO:0008026; F:ATP dependent helicase activity; IEA.		
DR	GO: GO:0005489; F:electron transporter activity; IEA.		
DR	GO: GO:0003723; F:RNA binding; IEA.		
DR	GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.		
DR	GO: GO:0008236; F:serine-type peptidase activity; IEA.		
DR	GO: GO:0005198; F:structural molecule activity; IEA.		
DR	GO: GO:0006118; F:electron transport; IEA.		
DR	GO: GO:0006508; F:proteolysis and peptidolysis; IEA.		
DR	GO: GO:0006350; F:transcription; IEA.		
DR	GO: GO:0019079; P:viral genome replication; IEA.		
DR	GO: GO:0019087; P:viral transformation; IEA.		
DR	Interpro: IPR009003; Cys Ser trypsin.		
DR	Interpro: IPR000345; CytC_heme_BS.		
DR	Interpro: IPR001410; DEAD.		
DR	Interpro: IPR002522; HCV_capsid.		
DR	Interpro: IPR002521; HCV_core.		
DR	Interpro: IPR002519; HCV_env.		
DR	Interpro: IPR002531; HCV_NS1.		
DR	Interpro: IPR002518; HCV_NS2.		
DR	Interpro: IPR000745; HCV_NS4a.		
DR	Interpro: IPR001490; HCV_NS4b.		
DR	Interpro: IPR002568; HCV_NS5a.		
DR	Interpro: IPR002166; HCV_NS5b.		
DR	Interpro: IPR001650; Helicase_C.		
DR	Interpro: IPR004109; Peptidase_C29.		
DR	Interpro: IPR007095; RNA_pol_DS_PS.		
DR	Interpro: IPR007094; RNA_pol_PSVir.		
DR	Pfam: PF01543; HCV_capsid; 1.		
DR	Pfam: PF01542; HCV_core; 1.		
DR	Pfam: PF01539; HCV_env; 1.		
DR	Pfam: PF01560; HCV_NS1; 1.		
DR	Pfam: PF01538; HCV_NS2; 1.		
DR	Pfam: PF02907; HCV_NS3; 1.		
DR	Pfam: PF01006; HCV_NS4a; 1.		
DR	Pfam: PF01001; HCV_NS4b; 1.		
DR	Pfam: PF01506; HCV_NS5a; 1.		
DR	Pfam: PF00271; helicase_C; 1.		
DR	Pfam: PF00998; Viral_RdRp; 1.		

DR PRODOM; ED186062; HCV NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KM Polypeptidein.
 SQ SEQUENCE 3010 AA; 327097 MW; EE6418C7A7235686 CRC64;
 Query Match 96.4%; Score 1775; DB 12; Length 3010;
 Best Local Similarity 95.2%; Pred. No. 1.8e-141;
 Matches 335; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLOWIPPLNVRGGRDAIILITCAVHPELIPDITKLLAIIPGLMVLQAGITKVPYFVR 60
 DB 855 AHLQMWVPLNIRGGRDAIILITCAVHPELIPDITKLLAIIPGLMVLQAGITKVPYFVR 914
 QY 61 AAGLIRACMLVYKRAAGHYVQMAFMKLAALTGTYYVYDHLPLQDMAHAGRLDAVAVEPV 120
 DB 915 AAGLIRACMLVYKRAAGHYVQMAFMKLAALTGTYYVYDHLPLQDMAHAGRLDAVAVEPV 974
 QY 121 IFSDMEXKIIITWGAADTAACGDIISGLPVSARGREIILGPAUNFEGGWLAPITAYSQ 180
 DB 975 VFSMEXKIIITWGAADTAACGDIISGLPVSARGREIILGPAUNFEGGWLAPITAYSQ 1034
 QY 181 QTRBLGCIITSLTGRDNQVVEGVVSTRATOSFLATCNVGVVTFHAGSKTTLAEPK 240
 DB 1035 QTRBLGCIITSLTGRDNQVVEGVVSTRATOSFLATCNVGVVTFHAGSKTTLAEPK 1094
 QY 241 GPIQMTNVDQDLVGMQAPGASMTPTCGSSDLYLVTRHADVIFVRRRSGSGLS 300
 DB 1095 GPIQMTNVDQDLVGMQAPGASMTPTCGSSDLYLVTRHADVIFVRRRSGSGLS 1154
 QY 301 PRPVSYIKGSSGGPLCPGSHAVGIFRAVCTRGVAVAVDFIPRESMTTMR 352
 DB 1155 PRPVSYIKGSSGGPLCPGSHAVGIFRAVCTRGVAVAVDFIPRESMTTMR 1206

RESULT 11
 ID 081755 PRELIMINARY; PRT; 1186 AA.
 AC 081755;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
 DS Polypeptidein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91089550; PubMed=2175503;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT "Molecular cloning of the human hepatitis C virus genome from Japanese
 RT patients with non-A, non-B hepatitis.";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
 RA Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RT J. Gen. Virol. 72:2697-2704(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91140698; PubMed=1847440;
 RA Takamizawa A., Mori C., Kanabe S., Murakami S., Fujita J., Onishi E.,
 RA Andoh T., Yoshida I., Okayama H.;
 RT "The structure and organization of the Hepatitis C virus genome
 RT isolated from human carriers.";
 RT J. Virol. 65:1105-1113(1991).

RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Seiby A., Barr P.J., Weiner A.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RT Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314445;
 RA Chen P., Lin M., Tai K., Lin P., Lin C., Chen D.;
 RT "The Taiwanese hepatitis C virus genome: Sequence determination and
 RT mapping the 5' terminus of viral genomic and antigenomic RNA.";
 RT Virology 188:102-113(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S., Yamamoto K., Iizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: Comparative study of four distinct
 RT genotypes.";
 RT Virology 188:331-341(1992).
 RN [7]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93323208; PubMed=8392606;
 RA Hijikata M., Mizushima H., Akagi T., Mori S., Kakuchi N., Kato N.,
 RA Tanaka T., Kimura K., Shimotohno K.;
 RT "Two distinct proteinase activities required for the processing of a
 RT putative nonstructural precursor protein of hepatitis C virus.";
 RT J. Virol. 67:4665-4675(1993).
 RN [8]
 RP SEQUENCE FROM N.A.
 RA Hijikata M.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D11397; BAA20975.1; -.
 DR PIR: A61196; A61196.
 DR PIR: P50329; P50329.
 DR PDB: 1DXP; 28-MAR-02.
 DR GO: GO:000524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0016787; F:hydrolyase activity; IEA.
 DR GO: GO:0003676; F:nucleic acid binding; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR001650; Helicase_C29.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW ATP-binding; Helicase; Hydrolase.
 FT NON TER
 SQ SEQUENCE 1186 AA; 126280 MW; 34170478BA23729A CRC64;

Query Match 96.3%; Score 1774; DB 12; Length 1186;
 Best Local Similarity 95.2%; Pred. No. 6.4e-142;
 Matches 335; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 1 AHLOWIPPLNVRGGRDAIILITCAVHPELIPDITKLLAIIPGLMVLQAGITKVPYFVR 60
 DB 133 AHLOWIPPLNVRGGRDAIILITCAVHPELIPDITKLLAIIPGLMVLQAGITKVPYFVR 192
 QY 61 AAGLIRACMLVYKRAAGHYVQMAFMKLAALTGTYYVYDHLPLQDMAHAGRLDAVAVEPV 120

```

Db 193 AAGLIRACMLVRKAGHYVOMAFMKLAALGTGVYDHLPLADMAHAGRDLAVAEPV 252
QY 121 IFSDEVEKIIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180
Db 253 VFSMEKRIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 312
QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSPLATCNVGMVTFHAGSKTLAGPK 240
Db 313 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSPLATCNVGMVTFHAGSKTLAGPK 372
QY 241 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSSDLYLVRHADVIIVRRRGDSRGSLLS 300
Db 373 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSSDLYLVRHADVIIVRRRGDSRGSLLS 432
QY 301 PRPVSYLKSSGGGGLCPSGHVAIGIFPAAVCTRGVAVAVDFIPVESMETTMR 352
Db 433 PRPVSYLKSSGGGGLCPSGHVAIGIFPAAVCTRGVAVAVDFIPVESMETTMR 484

RESULT 12
081817 PRELIMINARY; PRT; 2284 AA.
AC 081817
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Polypeptide precursor (Genome polypeptide).
OS Hepatitis C virus type 2.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=40271;
RN 11
RP SEQUENCE FROM N.A.
RA Hijioka M., Mitsuhashi H., Tanji Y., Komada Y., Hirowatari Y.,
RA Akagi T., Kimura K., Shimotohno K.,
RT "Proteolytic processing and membrane association of putative
RT nonstructural proteins of hepatitis C virus."
RL Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).
RN 12
RP SEQUENCE FROM N.A.
RA MEDLINE=9433810; Pubmed=8056334;
RA Tanji Y., Hijioka M., Hirowatari Y., Shimotohno K.,
RT "Identification of the domain required for trans-cleavage activity of
RT hepatitis C viral serine proteinase."
RL Gene 145:215-219(1994).
RN 13
RP SEQUENCE FROM N.A.
RA MEDLINE=95056078; Pubmed=7966638;
RA Tanji Y., Hijioka M., Hirowatari Y., Shimotohno K.,
RT "Hepatitis C virus polypeptide processing: kinetics and mutagenic
RT analysis of serine proteinase-dependent cleavage."
RL J. Virol. 68:8418-8422(1994).
RN 14
RP SEQUENCE FROM N.A.
RA MEDLINE=95156583; Pubmed=7853491;
RA Tanji Y., Hijioka M., Satoh S., Kaneko T., Shimotohno K.,
RT "Hepatitis C virus-encoded nonstructural protein NS4A has versatile
RT functions in viral protein processing."
RL J. Virol. 69:1575-1581(1995).
DR EMBL; D16435; BAA03905.1; -
DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; P00246; P00246.
DR PIR; P00246; P00246.
DR HSP; P26663; IXP.
DR GO; GO:0019012; C:viral; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003688; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008266; F:serine-type peptidase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

```

```

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR009003; Cys_Ser_Lypsin.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; peptidase_C29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVlr.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR Nonstructural protein; Polypeptide; RNA-directed RNA polymerase;
KW Signal; Transferase.
KM SIGNAL
FT SIGNAL 5 20
FT SIGNAL 5 83
FT CHAIN 21 54
FT CHAIN 84 300
FT CHAIN 301 931
FT CHAIN 932 985
FT CHAIN 986 1246
FT CHAIN 1247 1693
FT CHAIN 1694 2284
FT CHAIN NS5B.
SQ SEQUENCE 2284 AA; 247213 MW; DC272A1517046337 CRC64;

Query Match 96.3%; Score 1774; DB 12; Length 2284;
Best Local Similarity 95.2%; Pred. No. 1.5e-141;
Matches 335; Conservative 10; Mismatches 7; Indels 0; Gaps 0;
QY 1 AHLQWVPLPVRNRRGRDAIILLTCVHBEILFDITKLLAIFGLMVLQAGITVYFVR 60
Db 129 AHLQWVPLPVRNRRGRDAIILLTCVHBEILFDITKLLAIFGLMVLQAGITVYFVR 188
QY 61 AAGLIRACMLVRKAGHYVOMAFMKLAALGTGVYDHLPLADMAHAGRDLAVAEPV 120
Db 189 AAGLIRACMLVRKAGHYVOMAFMKLAALGTGVYDHLPLADMAHAGRDLAVAEPV 248
QY 121 IFSDEVEKIIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 180
Db 249 VFSDEKRIITWGADTAACDIIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 308
QY 181 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSPLATCNVGMVTFHAGSKTLAGPK 240
Db 309 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSPLATCNVGMVTFHAGSKTLAGPK 368
QY 241 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSSDLYLVRHADVIIVRRRGDSRGSLLS 300
Db 369 GPITOMYTNVDODLVGMQAPPGARSMTPTCTGSSSDLYLVRHADVIIVRRRGDSRGSLLS 428
QY 301 PRPVSYLKSSGGGGLCPSGHVAIGIFPAAVCTRGVAVAVDFIPVESMETTMR 352
Db 429 PRPVSYLKSSGGGGLCPSGHVAIGIFPAAVCTRGVAVAVDFIPVESMETTMR 480

RESULT 13
P89966 PRELIMINARY; PRT; 3010 AA.
AC P89966
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE RNA for polypeptide (Genome polypeptide).

```


OS Hepatitis C virus.
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=type 1b;
 RA Tanaka T.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=type 1b;
 RA TANAKA T.;
 RT "TMORF.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; D89872; BA14035.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00804; P00804.
 DR PIR; P50329; P50329.
 DR HSP; P26663; IUXP.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002531; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR000745; HCV NS4A.
 DR InterPro; IPR001490; HCV NS4B.
 DR InterPro; IPR002868; HCV NS5A.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVlr.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01543; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4A; 1.
 DR Pfam; PF01001; HCV NS4B; 1.
 DR Pfam; PF01506; HCV NS5A; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; viral_RdRp; 1.
 DR Pfam; PF01862; HCV NS1; 1.
 DR SMART; SM00467; DEXDC; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 337023 NW; E075BD9CFD8D1261 CRC64;

Matches 335; Conservative 10; Mismatches 7; Indels 0; Gaps 0;
 QY 1 AHLOWIPPLNVRGGRDAILITLCVAPPELLIDIKLLATFGLPMV;AAGTKVYFVR 60
 DB 855 AHLOWVPPPLNVRGGRDAILITLCVAPPELLIDIKLLATFGLPMV;AAGTKVYFVR 914
 QY 61 AAGLIRACMLVKAAGHYVQWAFMKLAALTGTVYDHLTPLDQWAAHGLRDLAAVEPV 120
 DB 915 AAGLIRACMLVKAAGHYVQWAFMKLAALTGTVYDHLTPLDQWAAHGLRDLAAVEPV 974
 QY 121 IFSDMEVKTITWGADTAACGDIISGLPVSARGREIILGPANFEGCGRLAPITAYSQ 180
 DB 975 VFSDMEVKTITWGADTAACGDIISGLPVSARGREIILGPANFEGCGRLAPITAYSQ 1034
 QY 181 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCTVFRGAGSKTLAEPK 240
 DB 1035 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSFLATCVNGVCTVFRGAGSKTLAEPK 1094
 QY 241 GPITQMTNVDDLVGMQAPGARSMPTCTGSSDLVYTRHADVIPIRRRDSRGLIS 300
 DB 1095 GPITQMTNVDDLVGMQAPGARSMPTCTGSSDLVYTRHADVIPIRRRDSRGLIS 1154
 QY 301 PRPVSTLKSGSGPILCPGSHAVGIFRAVCTRGVAKAVDFPVBSMETTKR 352
 DB 1155 PRPVSTLKSGSGPILCPGSHAVGIFRAVCTRGVAKAVDFPVBSMETTKR 1206
 RESULT 14
 ID Q9J3G6 PRELIMINARY; PRT: 3010 AA.
 AC Q9J3G6;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD26;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 RT progression.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AF207677; AA65957.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P00254; P00254.
 DR PIR; P50329; P50329.
 DR HSP; P26663; IUXP.
 DR MEROPS; S29.002; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0018028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.


```

DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV_NS4A.
DR InterPro: IPR001490; HCV_NS4B.
DR InterPro: IPR002868; HCV_NS5A.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR0041650; Helicase C.
DR InterPro: IPR004109; Peptidase C29.
DR InterPro: IPR007095; RNA_pol_D5_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4A; 1.
DR Pfam: PF01001; HCV_NS4B; 1.
DR Pfam: PF01506; HCV_NS5A; 1.
DR Pfam: PF00271; Helicase C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD16602; HCV_NS1; 1.
DR SMART: SM00487; DEXDC1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327165 MW; 74PAB6B80F24837B CRC64;

Query Match 96.3%; Score 1774; DB 12; Length 3010;
Best Local Similarity 95.5%; Pred. No. 2.1e-141;
Matches 336; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHLDWIPINVRGGRDAIILTCVAHPELIFDTIKLLAFGLPVLYLQAGTKVPEVR 60
DB 855 AHLDWIPINVRGGRDAIILTCVAHPELIFDTIKLLAFGLPVLYLQAGTKVPEVR 914
QY 61 AAGLIRACMLVRKAGGHVQMAFMKLAALGVYVVDHLLPLQDMAHAGRDVAVEPV 120
DB 915 AAGLIRACMLVRKAGGHVQMAFMKLAALGVYVVDHLLPLQDMAHAGRDVAVEPV 974
QY 121 IFSDMEXKITWAGDTAACGDIISGLPVSARSGREILGPDNFEQGWELLAPITAYSQ 180
DB 975 VFSMEXKITWAGDTAACGDIISGLPVSARSGREILGPDNFEQGWELLAPITAYSQ 1034
QY 181 QTRLLGCIITSLTGRKNOVEGEVQVSTATOSPLATCVNGVCMVYFHGASGKTLAGPK 240
DB 1035 QTRLLGCIITSLTGRKNOVEGEVQVSTATOSPLATCVNGVCMVYFHGASGKTLAGPK 1094
QY 241 GPITQWNTVNDODLVGQAPPGARSMPTCTGSSDLYLVTRHADVI PVRRSGDSRGLLS 300
DB 1095 GPITQWNTVNDODLVGQAPPGARSMPTCTGSSDLYLVTRHADVI PVRRSGDSRGLLS 1154
QY 301 PRPVSYKSGSGGFLCPGSHAVGIFRAAVCTRGAKAVDFIVESVETTMR 352
DB 1155 PRPVSYKSGSGGFLCPGSHAVGIFRAAVCTRGAKAVDFIVESVETTMR 1206

RESULT 15
Q99AU2 PRELIMINARY; PRT; 3010 AA.
Q99AU2
AC Q99AU2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus type 1b.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```

```

OC Hepatitis.
OX NCBI_TaxID=31647;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=chimera of HCV-BK;
RA Thomson M., Nascimbent M., Gonzales S., Murthy K., Rehmann B.,
  Liang J.;
RT "Analyses of viral sequences and virus-specific immune responses
  during serial passage of an infectious hepatitis C virus serotype 1b
  clone in chimpanzees."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND RNA (BY SIMILARITY).
CC EMBL: AF333324; AAK08509.1; -.
DR PIR: A61196; A61196.
DR PIR: P00246; P00246.
DR PIR: P00804; P00804.
DR PIR: P00329; P00329.
DR HSP: P26663; INB3.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F: electron transporter activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0016740; F: transferase activity; IEA.
DR GO: GO:0006118; P: electron transport; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P: transcription; IEA.
DR GO: GO:0019079; P: viral genome replication; IEA.
DR GO: GO:0019087; P: viral transformation; IEA.
DR InterPro: IPR009003; Cys Ser trypsin.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002518; HCV NS2.
DR InterPro: IPR000745; HCV_NS4A.
DR InterPro: IPR001490; HCV_NS4B.
DR InterPro: IPR002868; HCV_NS5A.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR0041650; Helicase C29.
DR InterPro: IPR004109; Peptidase C29.
DR InterPro: IPR007095; RNA_pol_D5_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4A; 1.
DR Pfam: PF01001; HCV_NS4B; 1.
DR Pfam: PF01506; HCV_NS5A; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD16602; HCV_NS1; 1.
DR SMART: SM00487; DEXDC1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 327007 MW; 053B9A653B0AB35 CRC64;

Query Match 96.3%; Score 1773; DB 12; Length 3010;
Best Local Similarity 94.6%; Pred. No. 2.6e-141;
Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13.8154 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-13
Perfect score: 1842
Sequence: 1 AHQWVIPPVLRGRDAI.....RGVAKAVDFPVESMETTMR 352

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1766	95.9	2201	4	US-09-539-601-6
2	1766	95.9	2201	4	US-09-539-601-15
3	1766	95.9	3010	4	US-09-539-601-3
4	1766	95.9	3010	4	US-09-539-601-21
5	1766	95.9	3010	4	US-09-539-601-27
6	1766	95.9	3010	4	US-09-539-601-33
7	1757	95.4	1692	3	US-09-263-933-4
8	1757	95.4	1692	3	US-09-263-933-2
9	1757	95.4	2307	3	US-09-263-933-2
10	1757	95.4	2307	3	US-09-263-933-11
11	1754	95.2	1692	3	US-09-263-933-11
12	1754	95.2	1692	3	US-09-263-933-9
13	1754	95.2	2307	3	US-09-263-933-9
14	1754	95.2	2307	3	US-09-263-933-9
15	1748	94.9	3010	3	US-09-014-416-3
16	1745	94.7	1692	3	US-09-263-933-18
17	1745	94.7	1692	3	US-09-263-933-18
18	1745	94.7	2307	3	US-09-263-933-16
19	1745	94.7	2307	3	US-09-263-933-16
20	1699	92.2	2013	2	US-08-324-977-12
21	1699	92.2	2013	2	US-08-324-977-12
22	1699	92.2	2013	2	US-08-324-977-12
23	1699	92.2	2013	2	US-08-324-977-12
24	1699	92.2	2013	2	US-08-324-977-12
25	1699	92.2	2013	2	US-08-324-977-12
26	1699	92.2	2013	2	US-08-324-977-12
27	1699	92.2	2013	2	US-08-324-977-12

28	1699	92.2	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1699	92.2	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1699	92.2	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1699	92.2	2621	2	US-08-384-616-36	Sequence 36, Appl
32	1699	92.2	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1699	92.2	3010	1	US-08-324-977-12	Sequence 2, Appl
34	1699	92.2	3010	1	US-08-324-977-12	Sequence 2, Appl
35	1699	92.2	3010	2	US-08-384-616-2	Sequence 14, Appl
36	1699	92.2	3010	2	US-08-384-616-14	Sequence 2, Appl
37	1699	92.2	3010	2	US-08-384-616-14	Sequence 2, Appl
38	1699	92.2	3010	2	US-08-384-616-14	Sequence 2, Appl
39	1699	92.2	3010	3	US-09-315-850-2	Sequence 14, Appl
40	1699	92.2	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1620	87.9	1648	5	US-08-188-281B-12	Sequence 12, Appl
42	1620	87.9	1648	5	PCT-US94-07280-12	Sequence 12, Appl
43	1620	87.9	1648	5	PCT-US95-01087-12	Sequence 12, Appl
44	1620	87.9	3011	1	US-08-188-281B-1	Sequence 1, Appl
45	1620	87.9	3011	1	US-08-453-552-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-539-601-6
; Sequence 6, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartschlag, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178-4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-539-601-6

Query Match 95.9%; Score 1766; DB 4; Length 2201;
Best Local Similarity 94.3%; Pred. No. 1.2e-170;
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY	1	1766	95.9	2201	4	US-09-539-601-6	Sequence 6, Appl
QY	2	1766	95.9	2201	4	US-09-539-601-15	Sequence 15, Appl
QY	3	1766	95.9	3010	4	US-09-539-601-3	Sequence 3, Appl
QY	4	1766	95.9	3010	4	US-09-539-601-21	Sequence 21, Appl
QY	5	1766	95.9	3010	4	US-09-539-601-27	Sequence 27, Appl
QY	6	1766	95.9	3010	4	US-09-539-601-33	Sequence 33, Appl
QY	7	1757	95.4	1692	3	US-09-263-933-4	Sequence 4, Appl
QY	8	1757	95.4	1692	3	US-09-263-933-2	Sequence 2, Appl
QY	9	1757	95.4	2307	3	US-09-263-933-2	Sequence 2, Appl
QY	10	1757	95.4	2307	3	US-09-263-933-11	Sequence 11, Appl
QY	11	1754	95.2	1692	3	US-09-263-933-11	Sequence 11, Appl
QY	12	1754	95.2	1692	3	US-09-263-933-9	Sequence 9, Appl
QY	13	1754	95.2	2307	3	US-09-263-933-9	Sequence 9, Appl
QY	14	1754	95.2	2307	3	US-09-263-933-9	Sequence 9, Appl
QY	15	1748	94.9	3010	3	US-09-014-416-3	Sequence 3, Appl
QY	16	1745	94.7	1692	3	US-09-263-933-18	Sequence 18, Appl
QY	17	1745	94.7	1692	3	US-09-263-933-18	Sequence 18, Appl
QY	18	1745	94.7	2307	3	US-09-263-933-16	Sequence 16, Appl
QY	19	1745	94.7	2307	3	US-09-263-933-16	Sequence 16, Appl
QY	20	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl
QY	21	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl
QY	22	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl
QY	23	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl
QY	24	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl
QY	25	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl
QY	26	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl
QY	27	1699	92.2	2013	2	US-08-324-977-12	Sequence 12, Appl

RESULT 2

US-09-539-601-15
 ; Sequence 15, Application US/09539601C
 ; Patent No. 6630343
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartschlagel, Ralf FW
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 ; FILE REFERENCE: all sequences
 ; CURRENT APPLICATION NUMBER: US/09/539,601C
 ; CURRENT FILING DATE: 2001-08-30
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 ; EARLIER FILING DATE: 1999-04-03
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 15
 ; LENGTH: 2201
 ; TYPE: PR1
 ; ORGANISM: Hepatitis C virus
 US-09-539-601-15

Query Match 95.9%; Score 1766; DB 4; Length 2201;
 Best Local Similarity 94.3%; Pred. No. 1.2e-170;
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60
 DB 46 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 105
 QY 61 AQLIRACMLVRKAGGYVQMAFMKLAALTGTYYVDTLPLODMAHAGRLDAVAEPV 120
 DB 106 AHGLIRACMLVRKAGGYVQMAFMKLAALTGTYYVDTLPLODMAHAGRLDAVAEPV 165
 QY 121 IFSMEVKIITWGDITACGDIISGLPVSARGREIILGPADNEGOGWRLAIPITAYSQ 180
 DB 166 VFSDEMETVITWGDITACGDIISGLPVSARGREIILGPADNEGOGWRLAIPITAYSQ 225
 QY 181 QTRGLGCIITSLGRDNQVEGVVSTATOSFLATCNVGVCTVPHGAGSKTLGPK 240
 DB 226 QTRGLGCIITSLGRDNQVEGVVSTATOSFLATCNVGVCTVPHGAGSKTLGPK 285
 QY 241 GPITOMTNNVDQIVGQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
 DB 286 GPITOMTNNVDQIVGQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 345
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 352
 DB 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 397

RESULT 3
 US-09-539-601-3
 ; Sequence 3, Application US/09539601C
 ; Patent No. 6630343
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartschlagel, Ralf FW
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 ; FILE REFERENCE: all sequences
 ; CURRENT APPLICATION NUMBER: US/09/539,601C
 ; CURRENT FILING DATE: 2001-08-30
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 ; EARLIER FILING DATE: 1999-04-03
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 3010
 ; TYPE: PR1
 ; ORGANISM: Hepatitis C virus
 US-09-539-601-3

Query Match 95.9%; Score 1766; DB 4; Length 3010;
 Best Local Similarity 94.3%; Pred. No. 1.9e-170;
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60

DB 855 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 914
 QY 61 AQLIRACMLVRKAGGYVQMAFMKLAALTGTYYVDTLPLODMAHAGRLDAVAEPV 120
 DB 915 AHGLIRACMLVRKAGGYVQMAFMKLAALTGTYYVDTLPLODMAHAGRLDAVAEPV 974
 QY 121 IFSMEVKIITWGDITACGDIISGLPVSARGREIILGPADNEGOGWRLAIPITAYSQ 180
 DB 975 VFSDEMETVITWGDITACGDIISGLPVSARGREIILGPADNEGOGWRLAIPITAYSQ 1034
 QY 181 QTRGLGCIITSLGRDNQVEGVVSTATOSFLATCNVGVCTVPHGAGSKTLGPK 240
 DB 1035 QTRGLGCIITSLGRDNQVEGVVSTATOSFLATCNVGVCTVPHGAGSKTLGPK 1094
 QY 241 GPITOMTNNVDQIVGQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
 DB 1095 GPITOMTNNVDQIVGQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 1154
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 352
 DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 1206

RESULT 4
 US-09-539-601-21
 ; Sequence 21, Application US/09539601C
 ; Patent No. 6630343
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartschlagel, Ralf FW
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 ; FILE REFERENCE: all sequences
 ; CURRENT APPLICATION NUMBER: US/09/539,601C
 ; CURRENT FILING DATE: 2001-08-30
 ; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 ; EARLIER FILING DATE: 1999-04-03
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 21
 ; LENGTH: 3010
 ; TYPE: PR1
 ; ORGANISM: Hepatitis C virus
 US-09-539-601-21

Query Match 95.9%; Score 1766; DB 4; Length 3010;
 Best Local Similarity 94.3%; Pred. No. 1.9e-170;
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60
 DB 855 AHLQWIPPLNVGRDAIILLTCVAHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 914
 QY 61 AQLIRACMLVRKAGGYVQMAFMKLAALTGTYYVDTLPLODMAHAGRLDAVAEPV 120
 DB 915 AHGLIRACMLVRKAGGYVQMAFMKLAALTGTYYVDTLPLODMAHAGRLDAVAEPV 974
 QY 121 IFSMEVKIITWGDITACGDIISGLPVSARGREIILGPADNEGOGWRLAIPITAYSQ 180
 DB 975 VFSDEMETVITWGDITACGDIISGLPVSARGREIILGPADNEGOGWRLAIPITAYSQ 1034
 QY 181 QTRGLGCIITSLGRDNQVEGVVSTATOSFLATCNVGVCTVPHGAGSKTLGPK 240
 DB 1035 QTRGLGCIITSLGRDNQVEGVVSTATOSFLATCNVGVCTVPHGAGSKTLGPK 1094
 QY 241 GPITOMTNNVDQIVGQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
 DB 1095 GPITOMTNNVDQIVGQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 1154
 QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 352
 DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETTMR 1206

```
RESULT 5
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartschlagel, Ralf FM
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; EARLIER FILING DATE: 2001-08-30
; EARLIER FILING DATE: 1991-15-178.4 GERMANY
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-27
```

```
Query Match 95.9%; Score 1766; DB 4; Length 3010;
Best Local Similarity 94.3%; Pred. No. 1.9e-170;
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
```

```
QY 1 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 60
DB 855 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 914
QY 61 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 120
DB 915 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 974
QY 121 IFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFGQWRLLAPITAYSQ 180
DB 975 VFSDEMTKIIITWGADTAACGDIISGLPVASRGRREILGPADNFGQWRLLAPITAYSQ 1034
QY 181 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCWTFVHAGSKTLAGPK 240
DB 1035 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCWTFVHAGSKTLAGPK 1094
QY 241 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIIVRRRDSRGLS 300
DB 1095 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIIVRRRDSRGLS 1154
QY 301 PRPVSYLKSGSGGGLLCPGSHAVGIFRAAVCTRGVAKAVDFVPSMETTMR 352
DB 1155 PRPVSYLKSGSGGGLLCPGSHAVGIFRAAVCTRGVAKAVDFVPSMETTMR 1206
```

```
RESULT 6
US-09-539-601-33
; Sequence 33, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartschlagel, Ralf FM
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; EARLIER FILING DATE: 2001-08-30
; EARLIER FILING DATE: 1991-15-178.4 GERMANY
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 33
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-33
```

```
Query Match 95.5%; Score 1759; DB 4; Length 3010;
Best Local Similarity 94.0%; Pred. No. 9.8e-170;
Matches 331; Conservative 11; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 1 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 60
DB 855 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 914
QY 61 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 120
DB 915 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 974
QY 121 IFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFGQWRLLAPITAYSQ 180
DB 975 VFSDEMTKIIITWGADTAACGDIISGLPVASRGRREILGPADNFGQWRLLAPITAYSQ 1034
QY 181 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCWTFVHAGSKTLAGPK 240
DB 1035 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCWTFVHAGSKTLAGPK 1094
QY 241 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIIVRRRDSRGLS 300
DB 1095 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIIVRRRDSRGLS 1154
QY 301 PRPVSYLKSGSGGGLLCPGSHAVGIFRAAVCTRGVAKAVDFVPSMETTMR 352
DB 1155 PRPVSYLKSGSGGGLLCPGSHAVGIFRAAVCTRGVAKAVDFVPSMETTMR 1206
```

```
RESULT 7
US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1999-03-08
; EARLIER FILING DATE: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-4
```

```
Query Match 95.4%; Score 1757; DB 3; Length 1692;
Best Local Similarity 93.8%; Pred. No. 6.7e-170;
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 1 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 60
DB 134 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 193
QY 61 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 120
DB 194 AHCQWIPPLNVRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVR 253
QY 121 IFSDMEVKIITWGADTAACGDIISGLPVASRGRREILGPADNFGQWRLLAPITAYSQ 180
DB 254 VFSDEMTKIIITWGADTAACGDIISGLPVASRGRREILGPADNFGQWRLLAPITAYSQ 313
QY 181 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCWTFVHAGSKTLAGPK 240
DB 314 QTRGLGCIITSLTGRDKQVEGEVQVSTATQSFATCVNGVCWTFVHAGSKTLAGPK 373
QY 241 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIIVRRRDSRGLS 300
DB 374 GPITQMTNVDDQDLVGMQAPPGARSMTPTCGSSDLVLTVRADVIIVRRRDSRGLS 433
QY 301 PRPVSYLKSGSGGGLLCPGSHAVGIFRAAVCTRGVAKAVDFVPSMETTMR 352
```

Db 434 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 485

RESULT 8

US-09-919-901-4
Sequence 4, Application US/09919901

Patent No. 6599738
GENERAL INFORMATION:

APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.

APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

FILE REFERENCE: 0125-0005A
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

CURRENT APPLICATION NUMBER: US/09/919, 901
CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 09/263, 933
PRIOR FILING DATE: 1999-02-08

PRIOR APPLICATION NUMBER: 09/129, 611
PRIOR FILING DATE: 1998-08-05

NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4
LENGTH: 1692

TYPE: PRT
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION:

US-09-919-901-4

Query Match

Best Local Similarity 95.4%; Score 1757; DB 4; Length 1692;
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

1 AHLQWIPPLNVRGGRDAIILTCVHPELFDITKLLAIFGPMVLAQGITKVPYFVR 60

134 AHLHWIPLNARGRDAIILMCAVHPELFDITKLLAIFGPMVLAQGITKVPYFVR 193

61 AAGLRACMLYRKAAGHYVQMAFMKLAALGTYYDHLTFLQWMAHAGLDLVAVEPV 120

194 AAGLIHACMLYRKAAGHYVQMAFMKLGALTGYIYNHLPLRDMAHAGLDLVAVEPV 253

121 IFSDEVKITITWGDITACGDIISGLPVSARKEIILGPADNIEGQWRLLAPITAYSQ 180

254 VFSDEVKITITWGDITACGDIISGLPVSARKEIILGPADNIEGQWRLLAPITAYSQ 313

181 QTRGLGCIITSLTGRDNKQVEGEVQVSTATQSFATCVNGVQWTVFHGASXTLAGPK 240

314 QTRGLGCIITSLTGRDNKQVEGEVQVSTATQSFATCVNGVQWTVFHGASXTLAGPK 373

241 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 300

374 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 433

301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

434 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 485

Db 241 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 300

374 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 433

301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

434 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 485

Db 241 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 300

374 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 433

301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

434 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 485

Db 241 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 300

374 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 433

301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

434 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 485

EARLIER APPLICATION NUMBER: 09/129, 611
EARLIER FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2307
TYPE: PRT
ORGANISM: Artificial Sequence
US-09-263-933-2

Query Match

Best Local Similarity 95.4%; Score 1757; DB 3; Length 2307;
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

1 AHLQWIPPLNVRGGRDAIILTCVHPELFDITKLLAIFGPMVLAQGITKVPYFVR 60

226 AHLHWIPLNARGRDAIILMCAVHPELFDITKLLAIFGPMVLAQGITKVPYFVR 285

61 AAGLRACMLYRKAAGHYVQMAFMKLAALGTYYDHLTFLQWMAHAGLDLVAVEPV 120

286 AAGLIHACMLYRKAAGHYVQMAFMKLGALTGYIYNHLPLRDMAHAGLDLVAVEPV 345

121 IFSDEVKITITWGDITACGDIISGLPVSARKEIILGPADNIEGQWRLLAPITAYSQ 180

346 VFSDEVKITITWGDITACGDIISGLPVSARKEIILGPADNIEGQWRLLAPITAYSQ 405

181 QTRGLGCIITSLTGRDNKQVEGEVQVSTATQSFATCVNGVQWTVFHGASXTLAGPK 240

406 QTRGLGCIITSLTGRDNKQVEGEVQVSTATQSFATCVNGVQWTVFHGASXTLAGPK 465

241 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 300

466 GPTQMTNVDQDILVGMQAPPGASMTPTCTGSSDLVLTFRHADVIFVRRGDSRGLLS 525

301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577

Db 301 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 352

526 PRPVSYLKSSGGPLCPSSGHAAGIFRAAVCTRGVAAVDFVPEVSEMTMR 577


```

QY 61 AAGIIRACMLVRKAGAGHYVQMAFMKLAALITGTYVYDHLTPLODMAGIRDLAAVEPV 120
DB 286 AAGIIRACMLVRKAGAGHYVQMAFMKLAALITGTYVYDHLTPLODMAGIRDLAAVEPV 345
QY 121 IFSDEMEVKIITWGDATACGDIISGLFVSARRGREIILGPADNEEGGWRLLAPITAYSQ 180
DB 346 VFSDEMEVKIITWGDATACGDIISGLFVSARRGREIILGPADNEEGGWRLLAPITAYSQ 405
QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCWTVFHHAGSKTLAGPK 240
DB 406 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCWTVFHHAGSKTLAGPK 465
QY 241 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTFRHADVIFVRRRDSRGSLS 300
DB 466 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTFRHADVIFVRRRDSRGSLS 525
QY 301 PRPVSYLKSGSGGPELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 352
DB 526 PRPVSYLKSGSGGPELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 577

```

```

RESULT 11
US-09-263-933-11
; Sequence 11, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potte, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-11

```

```

Query Match 95.2%; Score 1754; DB 3; Length 1692;
Best Local Similarity 93.5%; Pred. No. 1,4e-169;
Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 AHLQWIPPLNVRGGRDAIILITCAVHPELIDITKLLAIPEGPLVYLQGITKVPYFVR 60
DB 134 AHLQWIPPLNVRGGRDAIILITCAVHPELIDITKLLAIPEGPLVYLQGITKVPYFVR 193
QY 61 AAGIIRACMLVRKAGAGHYVQMAFMKLAALITGTYVYDHLTPLODMAGIRDLAAVEPV 120
DB 194 AAGIIRACMLVRKAGAGHYVQMAFMKLAALITGTYVYDHLTPLODMAGIRDLAAVEPV 253
QY 121 IFSDEMEVKIITWGDATACGDIISGLFVSARRGREIILGPADNEEGGWRLLAPITAYSQ 180
DB 254 VFSDEMEVKIITWGDATACGDIISGLFVSARRGREIILGPADNEEGGWRLLAPITAYSQ 313
QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCWTVFHHAGSKTLAGPK 240
DB 314 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCWTVFHHAGSKTLAGPK 373
QY 241 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTFRHADVIFVRRRDSRGSLS 300
DB 374 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTFRHADVIFVRRRDSRGSLS 433
QY 301 PRPVSYLKSGSGGPELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 352
DB 434 PRPVSYLKSGSGGPELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 485

```

```

RESULT 12
US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potte, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-11

```

```

Query Match 95.2%; Score 1754; DB 4; Length 1692;
Best Local Similarity 93.5%; Pred. No. 1,4e-169;
Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 AHLQWIPPLNVRGGRDAIILITCAVHPELIDITKLLAIPEGPLVYLQGITKVPYFVR 60
DB 134 AHLQWIPPLNVRGGRDAIILITCAVHPELIDITKLLAIPEGPLVYLQGITKVPYFVR 193
QY 61 AAGIIRACMLVRKAGAGHYVQMAFMKLAALITGTYVYDHLTPLODMAGIRDLAAVEPV 120
DB 194 AAGIIRACMLVRKAGAGHYVQMAFMKLAALITGTYVYDHLTPLODMAGIRDLAAVEPV 253
QY 121 IFSDEMEVKIITWGDATACGDIISGLFVSARRGREIILGPADNEEGGWRLLAPITAYSQ 180
DB 254 VFSDEMEVKIITWGDATACGDIISGLFVSARRGREIILGPADNEEGGWRLLAPITAYSQ 313
QY 181 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCWTVFHHAGSKTLAGPK 240
DB 314 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGVCWTVFHHAGSKTLAGPK 373
QY 241 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTFRHADVIFVRRRDSRGSLS 300
DB 374 GPITQWYTNVDODLVGMQAPPGARSMTPTCGSSDLVLTFRHADVIFVRRRDSRGSLS 433
QY 301 PRPVSYLKSGSGGPELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 352
DB 434 PRPVSYLKSGSGGPELCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 485

```

```

RESULT 13
US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potte, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33

```

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 2307
TYPE: PRF
ORGANISM: Artificial Sequence
US-09-263-933-9

Query Match 95.2%; Score 1754; DB 3; Length 2307;
Best Local Similarity 93.5%; Pred. No. 2.1e-169;
Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHHQVWIPPLNVGRGDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 60
DB 226 AHHQVWIPPLNVGRGDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 285
QY 61 AAGLRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 120
DB 286 AAGLRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 345
QY 121 IFSDMEVKIITWGDITACDIIISGLPVSARGRBILGPADNFEQGMRLAPTAYSQ 180
DB 346 VFSDMETKIIITWGDITACDIIISGLPVSARGRBILGPADNFEQGMRLAPTAYSQ 405
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 240
DB 406 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 465
QY 241 GPITOMYTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
DB 466 GPITOMYTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 525
QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAADVPIVPSMETTMR 352
DB 526 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAADVPIVPSMETTMR 577

RESULT 14

US-09-919-901-9
Sequence 9, Application US/09919901
Patent No. 6599738
GENERAL INFORMATION:
APPLICANT: Potes, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/919, 901
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263, 933
PRIOR FILING DATE: 1999-02-08
PRIOR APPLICATION NUMBER: 09/129, 611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 2307
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: :
US-09-919-901-9

Query Match 95.2%; Score 1754; DB 4; Length 2307;
Best Local Similarity 93.5%; Pred. No. 2.1e-169;
Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHHQVWIPPLNVGRGDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 60
DB 226 AHHQVWIPPLNVGRGDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 285
QY 61 AAGLRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 120

DB 286 AAGLRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 345
QY 121 IFSDMEVKIITWGDITACDIIISGLPVSARGRBILGPADNFEQGMRLAPTAYSQ 180
DB 346 VFSDMETKIIITWGDITACDIIISGLPVSARGRBILGPADNFEQGMRLAPTAYSQ 405
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 240
DB 406 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 465
QY 241 GPITOMYTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
DB 466 GPITOMYTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 525
QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAADVPIVPSMETTMR 352
DB 526 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAADVPIVPSMETTMR 577

RESULT 15

US-09-014-416-3
Sequence 3, Application US/09014416
Patent No. 6153421
GENERAL INFORMATION:
APPLICANT: Yanagi, Masaaki
APPLICANT: Bukh, Jens
APPLICANT: Emerson, Suzanne U.
APPLICANT: Purcell, Robert H.
TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
FILE REFERENCE: 20264276
CURRENT APPLICATION NUMBER: US/09/014, 416
CURRENT FILING DATE: 1998-01-27
EARLIER FILING DATE: 1997-07-18
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 3010
TYPE: PRF
ORGANISM: Hepatitis C virus
US-09-014-416-3

Query Match 94.9%; Score 1748; DB 3; Length 3010;
Best Local Similarity 92.9%; Pred. No. 1.3e-168;
Matches 327; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 AHHQVWIPPLNVGRGDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 60
DB 855 AHHQVWIPPLNVGRGDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVR 914
QY 61 AAGLRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 120
DB 915 AAGLRACMLVRKKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 121 IFSDMEVKIITWGDITACDIIISGLPVSARGRBILGPADNFEQGMRLAPTAYSQ 180
DB 975 VFSAMETVITWGDITACDIIISGLPVSARGRBILGPADNFEQGMRLAPTAYSQ 1034
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 240
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSXTLAGPK 1094
QY 241 GPITOMYTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
DB 1095 GPITOMYTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 1154
QY 301 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAADVPIVPSMETTMR 352
DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAADVPIVPSMETTMR 1206

Search completed: May 6, 2004, 09:39:04

Fri May 7 13:37:03 2004

Job time : 14.8154 secs

us-10-650-585-13.ra1

Page 7

Db 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240
 QY 241 GPTQMTNVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300
 Db 241 GPTQMTNVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300
 QY 301 PRPVSYLKSSGGPPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352
 Db 301 PRPVSYLKSSGGPPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352

RESULT 2

US-10-650-585-13
 ; Sequence 13, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 352
 ; TYPE: PRT
 ; ORGANISM: HCV
 ; US-10-650-585-13

Query Match 100.0%; Score 1842; DB 16; Length 352;
 Best Local Similarity 100.0%; Pred. No. 5,7e-180;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHQVWIPPLNVGRDAIILLTCAVHPELIFDITKLLAIFGFLMVLQAGITVPEFVR 60
 Db 1 AHQVWIPPLNVGRDAIILLTCAVHPELIFDITKLLAIFGFLMVLQAGITVPEFVR 60
 QY 61 AOGIRACMLVRKAGGYVQMAFMKLAALGTGVYDHLTPLODMAHAGRLDAVAEPV 120
 Db 61 AOGIRACMLVRKAGGYVQMAFMKLAALGTGVYDHLTPLODMAHAGRLDAVAEPV 120
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGWRLAFTAYSQ 180
 Db 121 IFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGWRLAFTAYSQ 180
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240
 Db 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240
 QY 241 GPTQMTNVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300
 Db 241 GPTQMTNVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300
 QY 301 PRPVSYLKSSGGPPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352
 Db 301 PRPVSYLKSSGGPPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352

RESULT 3

US-10-017-736-12
 ; Sequence 12, Application US/10017736
 ; Publication No. US20020192640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/017,736
 ; CURRENT FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 380
 ; TYPE: PRT
 ; ORGANISM: HCV
 ; US-10-017-736-12

Query Match 100.0%; Score 1842; DB 13; Length 380;
 Best Local Similarity 100.0%; Pred. No. 6,4e-180;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHQVWIPPLNVGRDAIILLTCAVHPELIFDITKLLAIFGFLMVLQAGITVPEFVR 60
 Db 29 AHQVWIPPLNVGRDAIILLTCAVHPELIFDITKLLAIFGFLMVLQAGITVPEFVR 86
 QY 61 AOGIRACMLVRKAGGYVQMAFMKLAALGTGVYDHLTPLODMAHAGRLDAVAEPV 120
 Db 89 AOGIRACMLVRKAGGYVQMAFMKLAALGTGVYDHLTPLODMAHAGRLDAVAEPV 148
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGWRLAFTAYSQ 180
 Db 149 IFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGWRLAFTAYSQ 208
 QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 240
 Db 209 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPX 268
 QY 241 GPTQMTNVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 300
 Db 269 GPTQMTNVDQDLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGLLS 328
 QY 301 PRPVSYLKSSGGPPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 352
 Db 329 PRPVSYLKSSGGPPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETMR 380

RESULT 4

US-10-650-585-12
 ; Sequence 12, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 380
 ; TYPE: PRT
 ; ORGANISM: HCV
 ; US-10-650-585-12

Query Match 100.0%; Score 1842; DB 16; Length 380;
 Best Local Similarity 100.0%; Pred. No. 6,4e-180;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHQVWIPPLNVGRDAIILLTCAVHPELIFDITKLLAIFGFLMVLQAGITVPEFVR 60
 Db 29 AHQVWIPPLNVGRDAIILLTCAVHPELIFDITKLLAIFGFLMVLQAGITVPEFVR 86
 QY 61 AOGIRACMLVRKAGGYVQMAFMKLAALGTGVYDHLTPLODMAHAGRLDAVAEPV 120
 Db 89 AOGIRACMLVRKAGGYVQMAFMKLAALGTGVYDHLTPLODMAHAGRLDAVAEPV 148
 QY 121 IFSDMEVKIITWGADTAACGDIISGLPVSAARGREILLGPADNFEQGWRLAFTAYSQ 180

```

Db 149 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGPADNFEQGMWLLAPITAYSQ 208
QY 181 QTRGLGCIITSLTGRDKQVEGEVOVSTATOSFLATCNVCWTFHAGSKTLAAGPK 240
Db 209 QTRGLGCIITSLTGRDKQVEGEVOVSTATOSFLATCNVCWTFHAGSKTLAAGPK 268
QY 241 GPITQMTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
Db 269 GPITQMTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 328
QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMETMR 352
Db 329 PRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMETMR 380

```

RESULT 5

```

US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

```

```

Query Match 100.0%; Score 1842; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,7e-180;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AHQWIPPLNVRGSDATIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 60
Db 42 AHQWIPPLNVRGSDATIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 101
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAVEPV 120
Db 102 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAVEPV 161
QY 121 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGPADNFEQGMWLLAPITAYSQ 180
Db 162 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGPADNFEQGMWLLAPITAYSQ 221
QY 181 QTRGLGCIITSLTGRDKQVEGEVOVSTATOSFLATCNVCWTFHAGSKTLAAGPK 240
Db 222 QTRGLGCIITSLTGRDKQVEGEVOVSTATOSFLATCNVCWTFHAGSKTLAAGPK 281
QY 241 GPITQMTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
Db 282 GPITQMTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 341
QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMETMR 352
Db 342 PRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMETMR 393

```

RESULT 6

```

US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082

```

```

; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

```

```

Query Match 100.0%; Score 1842; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 6,7e-180;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AHQWIPPLNVRGSDATIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 60
Db 42 AHQWIPPLNVRGSDATIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 101
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAVEPV 120
Db 102 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAVEPV 161
QY 121 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGPADNFEQGMWLLAPITAYSQ 180
Db 162 IFSMEVKIITWGDADTAACGDIISGLPVSARREIILGPADNFEQGMWLLAPITAYSQ 221
QY 181 QTRGLGCIITSLTGRDKQVEGEVOVSTATOSFLATCNVCWTFHAGSKTLAAGPK 240
Db 222 QTRGLGCIITSLTGRDKQVEGEVOVSTATOSFLATCNVCWTFHAGSKTLAAGPK 281
QY 241 GPITQMTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
Db 282 GPITQMTNVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 341
QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMETMR 352
Db 342 PRPVSYLKSSGGPILCPSGHAGVIFRAVCTRGVAKAVDFIPVESMETMR 393

```

RESULT 7

```

US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2

```

```

Query Match 100.0%; Score 1842; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 7,1e-180;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AHQWIPPLNVRGSDATIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 60
Db 46 AHQWIPPLNVRGSDATIIILTCVHPELFDITKLLAIFGRLMVLQAGITKVPYFVR 105
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAVEPV 120

```


Db 106 AAGIIRACMTVRKAGGHYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 165
Qy 121 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILGPDNPEQGMRLAPITAYQ 180
Db 166 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILGPDNPEQGMRLAPITAYQ 225
Qy 181 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFPHGAGSKTLAGPK 240
Db 226 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFPHGAGSKTLAGPK 285
Qy 241 GPTQMTNTVDQDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGLS 300
Db 286 GPTQMTNTVDQDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGLS 345
Qy 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 352
Db 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 397

RESULT 8
US-10-650-585-2

Sequence 2, Application US/10650585
Publication No. US20040077066A1
GENERAL INFORMATION:
APPLICANT: Boehringer Ingelheim (Canada) Ltd.
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
FILE REFERENCE: 13/082
CURRENT APPLICATION NUMBER: US/10/650,585
CURRENT FILING DATE: 2003-08-28
PRIOR APPLICATION NUMBER: US/10/017,736A
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: 60/256,031
PRIOR FILING DATE: 2000-12-15
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 409
TYPE: PRT
ORGANISM: HCV
US-10-650-585-2

Query Match 100.0%; Score 1842; DB 16; Length 409;
Best Local Similarity 100.0%; Pred. No. 7,1e-180;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHQVWIPPLNVGRGRDAIILLTCVAHPLELFDITKLLAIFGLMWLOAGITKVEYFVR 60
Db 46 AHQVWIPPLNVGRGRDAIILLTCVAHPLELFDITKLLAIFGLMWLOAGITKVEYFVR 105
Qy 61 AAGIIRACMTVRKAGGHYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 120
Db 106 AAGIIRACMTVRKAGGHYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPV 165
Qy 121 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILGPDNPEQGMRLAPITAYQ 180
Db 166 IFSMEVKIITWGADTAACGDIISGLPVSARRGREILGPDNPEQGMRLAPITAYQ 225
Qy 181 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFPHGAGSKTLAGPK 240
Db 226 QTRGLACIITSLTGRKNOVEGEVQVSTATOSFLATCNVCWTFPHGAGSKTLAGPK 285
Qy 241 GPTQMTNTVDQDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGLS 300
Db 286 GPTQMTNTVDQDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGLS 345
Qy 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 352
Db 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 397

RESULT 9
US-10-017-736-14
Sequence 14, Application US/10017736

Publication No. US20020192640A1
GENERAL INFORMATION:
APPLICANT: Boehringer Ingelheim (Canada) Ltd.
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
FILE REFERENCE: 13/082
CURRENT APPLICATION NUMBER: US/10/017,736
CURRENT FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: 60/256,031
PRIOR FILING DATE: 2000-12-15
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 341
TYPE: PRT
ORGANISM: HCV
US-10-017-736-14

Query Match 96.5%; Score 1778; DB 13; Length 341;
Best Local Similarity 100.0%; Pred. No. 2e-173;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 VRGGRDAIILLTCVAHPLELFDITKLLAIFGLMWLOAGITKVEYFVRAGILRACMLV 71
Db 1 VRGGRDAIILLTCVAHPLELFDITKLLAIFGLMWLOAGITKVEYFVRAGILRACMLV 60
Qy 72 RRAAGGHYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPVIFSMEVKIIT 131
Db 61 RRAAGGHYQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAVEPVIFSMEVKIIT 120
Qy 122 WGADTAACGDIISGLPVSARRGREILGPDNPEQGMRLAPITAYSQTRGLACIIT 191
Db 121 WGADTAACGDIISGLPVSARRGREILGPDNPEQGMRLAPITAYSQTRGLACIIT 180
Qy 192 SLTGRDKNOVEGEVQVSTATOSFLATCNVCWTFPHGAGSKTLAGPKGPTQMTNTVD 251
Db 181 SLTGRDKNOVEGEVQVSTATOSFLATCNVCWTFPHGAGSKTLAGPKGPTQMTNTVD 240
Qy 252 QDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGLSPRPVSYLKSS 311
Db 241 QDLVGMQAPPGARSMTPTCGSSDLYLVRHADVIPIVRRGDSRGLSPRPVSYLKSS 300
Qy 312 GGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 352
Db 301 GGPLCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTR 341

RESULT 10
US-10-650-585-14
Sequence 14, Application US/10650585
Publication No. US20040077066A1
GENERAL INFORMATION:
APPLICANT: Boehringer Ingelheim (Canada) Ltd.
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
FILE REFERENCE: 13/082
CURRENT APPLICATION NUMBER: US/10/650,585
CURRENT FILING DATE: 2003-08-28
PRIOR APPLICATION NUMBER: US/10/017,736A
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: 60/256,031
PRIOR FILING DATE: 2000-12-15
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 341
TYPE: PRT
ORGANISM: HCV
US-10-650-585-14

Query Match 96.5%; Score 1778; DB 16; Length 341;
Best Local Similarity 100.0%; Pred. No. 2e-173;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 VRGGRDAIILLTCVAHPLELFDITKLLAIFGLMWLOAGITKVEYFVRAGILRACMLV 71

```

Db 1 VRGRDAIILITCAVHELFIDITKLLALFGLPLMTVOAGITKVPYVRAQGLIRACMLV 60
Qy 72 RKAAGHYVQMAFMKLAALITGYTYYDLTPLODMAHAGLDLAVAVEPVIFSMETKITT 131
Db 61 RKAAGHYVQMAFMKLAALITGYTYYDLTPLODMAHAGLDLAVAVEPVIFSMETKITT 120
Qy 132 WGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSOOTRGLGCIIT 191
Db 121 WGAADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSOOTRGLGCIIT 180
Qy 192 SLTRDKNQVGEVQVSTATOSFLATCVNGVCTVFGAGSKTLAPKGPITOMYTNVD 251
Db 181 SLTRDKNQVGEVQVSTATOSFLATCVNGVCTVFGAGSKTLAPKGPITOMYTNVD 240
Qy 252 QDLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPVSYLKGSS 311
Db 241 QDLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIPIVRRRDSRGSILSPVSYLKGSS 300
Qy 312 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTWR 352
Db 301 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTWR 341

```

RESULT 11

```

US-10-029-907-3
; Sequence 3, Application US/10023907
; Publication No. US20020142350A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029,907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
US-10-029-907-3

```

```

Query Match 95.9%; Score 1766; DB 13; Length 2201;
Best Local Similarity 94.3%; Pred. No. 4,7e-171;
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

```

Qy 1 AHLQWIPPLNVRGGRDAIILITCAVHELFIDITKLLALFGLPLMTVOAGITKVPYFVR 60
Db 46 AHLQWIPPLNVRGGRDAIILITCAIHPELIFTIKLLALIGPLMTVOAGITKVPYFVR 105
Qy 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALITGYTYYDLTPLODMAHAGLDLAVAVEPV 120
Db 106 AAGLIRACMLVRKAAGHYVQMAFMKLAALITGYTYYDLTPLODMAHAGLDLAVAVEPV 165
Qy 121 IFSMEVKITTWGADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSQ 180
Db 166 VFSMEVKITTWGADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSQ 225
Qy 181 QTRGLGCIITSLTRDKNQVGEVQVSTATOSFLATCVNGVCTVFGAGSKTLAPK 240
Db 226 QTRGLGCIITSLTRDKNQVGEVQVSTATOSFLATCVNGVCTVFGAGSKTLAPK 285
Qy 241 GPITOMYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIPIVRRRDSRGSILS 300

```

```

Db 286 GPITOMYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIPIVRRRDSRGSILS 345
Qy 301 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTWR 352
Db 346 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTWR 397

```

RESULT 12

```

US-10-309-561-3
; Sequence 3, Application US/10309561
; Publication No. US20030148348A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/309,561
; PRIOR FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US/10/029,907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,857
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
US-10-309-561-3

```

```

Query Match 95.9%; Score 1766; DB 14; Length 2201;
Best Local Similarity 94.3%; Pred. No. 4,7e-171;
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

```

Qy 1 AHLQWIPPLNVRGGRDAIILITCAVHELFIDITKLLALFGLPLMTVOAGITKVPYFVR 60
Db 46 AHLQWIPPLNVRGGRDAIILITCAIHPELIFTIKLLALIGPLMTVOAGITKVPYFVR 105
Qy 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALITGYTYYDLTPLODMAHAGLDLAVAVEPV 120
Db 106 AAGLIRACMLVRKAAGHYVQMAFMKLAALITGYTYYDLTPLODMAHAGLDLAVAVEPV 165
Qy 121 IFSMEVKITTWGADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSQ 180
Db 166 VFSMEVKITTWGADTAACGDIISGLPVASARREIILGPADNFEQGMRLIAPITAYSQ 225
Qy 181 QTRGLGCIITSLTRDKNQVGEVQVSTATOSFLATCVNGVCTVFGAGSKTLAPK 240
Db 226 QTRGLGCIITSLTRDKNQVGEVQVSTATOSFLATCVNGVCTVFGAGSKTLAPK 285
Qy 241 GPITOMYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIPIVRRRDSRGSILS 300
Db 286 GPITOMYTNVDQDLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIPIVRRRDSRGSILS 345
Qy 301 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTWR 352
Db 346 PRPVSYLKGSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDFPVESMETTWR 397

```

RESULT 13

```

US-10-467-000-1
; Sequence 1, Application US/10467000
; Publication No. US20040067486A1
; GENERAL INFORMATION:

```

```

; APPLICANT: De Francesco, Raffaele
; APPLICANT: Magliacchio, Giovanni
; APPLICANT: Pavesa, Giacomo
; TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
; FILE OF INVENTION: ENHANCED CELLS
; FILE REFERENCE: ITR0003P
; CURRENT APPLICATION NUMBER: US/10/467,000
; PRIOR FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: PCT/EP02/00526
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/263,479
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Con 1 HCV isolate nucleic acid
US-10-467-000-1

Query Match          95.4%; Score 1766; DB 12; Length 3010;
Best Local Similarity 94.3%; Pred. No. 7,4e-171;
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHHQWIPPLNVRGGRAIILITCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60
DB AHHQWIPPLNVRGGRAIILITCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 914
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 120
DB AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 974
QY 915 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 974
DB AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 974

QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 180
DB 975 VFSMEKVIITWGAADTAACDIIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 1034
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVQWTVFRRGDSRGLS 240
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVQWTVFRRGDSRGLS 1094
QY 241 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVTSHADVIPIVRRGDSRGLS 300
DB 1095 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVTSHADVIPIVRRGDSRGLS 1154
QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESNETTMR 352
DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESNETTMR 1206

RESULT 14
US-09-919-901-4
; Sequence 4, Application US/09919901
; Publication No. US20030082518A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; OTHER INFORMATION:
US-09-919-901-4

Query Match          95.4%; Score 1757; DB 10; Length 1692;
Best Local Similarity 93.8%; Pred. No. 2.7e-170;
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHHQWIPPLNVRGGRAIILITCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60
DB 134 AHHQWIPPLNVRGGRAIILITCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 193
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 120
DB 194 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 253
QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 180
DB 254 VFSMEKVIITWGAADTAACDIIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 313
QY 181 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVQWTVFRRGDSRGLS 240
DB 314 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCNGVQWTVFRRGDSRGLS 373
QY 241 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVTSHADVIPIVRRGDSRGLS 300
DB 374 GPITQMTNVQDILVGMQAPPGARSMTPTCGSSDLYLVTSHADVIPIVRRGDSRGLS 433
QY 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESNETTMR 352
DB 434 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKADVPIVESNETTMR 485

RESULT 15
US-10-191-966-4
; Sequence 4, Application US/10191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/263,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-10-191-966-4

Query Match          95.4%; Score 1757; DB 14; Length 1692;
Best Local Similarity 93.8%; Pred. No. 2.7e-170;
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHHQWIPPLNVRGGRAIILITCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 60
DB 134 AHHQWIPPLNVRGGRAIILITCAVHPELIFDITKLLAIFGLMVLQAGITKVPYFVR 193
QY 61 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 120
DB 194 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 253
QY 121 IFSDMEVKIITWGAADTAACDIIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 180
DB 194 AAGLIRACMLVRKAAGHYVQMAFMKLAALTGTYYVDHLTFLDWMHAGRLDAVAEVPV 253

```

```

Db      254 VESDMETKITWGAADTAACDIIILGPVSARRGKEIILGPADSIEGRGMRLAPITAYSQ 313
QY      181 QTRGLGCIITSLTGRDKNQVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 240
          |||||
Db      314 QTRGLGCIITSLTGRDKNQVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 373
          |||||
QY      241 GPITOMTNTVDQDIVGMOAPPGARSMTPCTGSSDLVLTFRHADVIPIVRRRGDSRGSLLS 300
          |||||
Db      374 GPITOMTNTVDQDIVGMOAPPGARSLTPCTGSSDLVLTFRHADVIPIVRRRGDSRGSLLS 433
          |||||
QY      301 PRPVSYLKSSSGGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETMR 352
          |||||
Db      434 PRPVSYLKSSSGGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETMR 485
          |||||

```

Search completed: May 6, 2004, 09:43:21
 Job time : 37.0084 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 46.4159 Seconds
(without alignments)
2075.771 Million cell updates/sec

Title: US-10-650-585-14
Perfect score: 1778
Sequence: 1 VRRGRDAIILLTCVAPHELI.....RGVAKAVDFPVSMTETWR 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1778	100.0	341	5	ABG32187 HCV prote
2	1778	100.0	352	5	ABG32186 HCV prote
3	1778	100.0	380	5	ABG32185 HCV prote
4	1778	100.0	393	5	ABG32184 HCV prote
5	1778	100.0	409	5	ABG32181 HCV prote
6	1710	96.2	3010	2	AAR82694 Partial H
7	1705	95.9	3010	2	AAR82694 HCV prote
8	1704	95.8	3010	2	AAR82694 HCV prote
9	1702	95.7	2201	5	ABG30591 Hepatitis
10	1702	95.7	2201	5	ABG30591 Hepatitis
11	1702	95.7	2201	5	ABG30591 Hepatitis
12	1702	95.7	2201	5	ABG30591 Hepatitis
13	1702	95.7	2201	5	ABG30591 Hepatitis
14	1702	95.7	2201	5	ABG30591 Hepatitis
15	1702	95.7	2201	5	ABG30591 Hepatitis
16	1702	95.7	2201	5	ABG30591 Hepatitis
17	1702	95.7	2201	5	ABG30591 Hepatitis
18	1702	95.7	2201	5	ABG30591 Hepatitis
19	1702	95.7	2201	5	ABG30591 Hepatitis
20	1702	95.7	2201	5	ABG30591 Hepatitis
21	1702	95.7	2201	5	ABG30591 Hepatitis
22	1702	95.7	2201	5	ABG30591 Hepatitis
23	1702	95.7	2201	5	ABG30591 Hepatitis
24	1702	95.7	2201	5	ABG30591 Hepatitis
25	1702	95.7	2201	5	ABG30591 Hepatitis

26	1702	95.7	3010	5	ABG32460 Hepatitis
27	1702	95.7	3010	5	ABG32461 Hepatitis
28	1702	95.7	3010	5	ABG32454 Hepatitis
29	1702	95.7	3011	5	ABG32456 Hepatitis
30	1699	95.6	2201	5	ABG30586 Hepatitis
31	1699	95.6	2201	5	ABG30589 Hepatitis
32	1699	95.6	2201	5	ABG30583 Hepatitis
33	1699	95.6	2201	5	ABG30588 Hepatitis
34	1698	95.5	2201	5	ABG30590 Hepatitis
35	1698	95.5	2307	3	AA770064 Recombina
36	1697	95.4	3010	5	AAE20477 HCV-S1 fu
37	1696	95.4	3010	5	ABG32452 Recombina
38	1695	95.3	768	2	AAR40223 Recombina
39	1695	95.3	2201	5	ABG30584 Hepatitis
40	1695	95.3	2201	5	ABG30602 Recombina
41	1695	95.3	2307	3	AA770065 Recombina
42	1695	95.3	3010	5	ABG32453 Hepatitis
43	1691	95.1	2201	5	ABG30585 Hepatitis
44	1688	94.9	3014	2	AAR54099 NANBV E1
45	1687	94.9	3010	2	AAW98022 Infectiou

ALIGNMENTS

RESULT 1	ABG32187	standard; protein; 341 AA.
ID	ABG32187	
XX	ABG32187	
AC	ABG32187	
XX		
DT	05-NOV-2002	(first entry)
XX		
DE	HCV protease NS2/3 truncation mutant 866-1206.	
XX		
KM	HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;	
KM	chronic liver disease; cirrhosis; end-stage liver disease; viraemia;	
KM	hepatocellular; antiinflammatory; lauryldiethylamine oxide; LDAO;	
KM	chaotropic agent; mutant; mutain.	
XX		
OS	Hepatitis C virus.	
OS	Synthetic.	
PN	W0200248375-A2.	
XX		
PD	20-UTN-2002.	
XX		
PF	13-DEC-2001; 2001WO-CA001796.	
XX		
PR	15-DEC-2000; 2000US-0256031P.	
PA	(BOEH) BOEHRINGER INGELHEIM CANADA LTD.	
XX		
PI	Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;	
XX	WPI; 2002-599511/64.	
DR		
XX		
PT	Novel polypeptide for screening inhibitors of non-structural proteases	
PT	useful as therapeutic agents against hepatitis C virus, comprises full	
PT	length non-structural protease, or its truncation.	
XX		
PS	Claim 41; Page 62-63; 67pp; English.	
CC	The invention relates to an isolated polypeptide consisting of a full-	
CC	length HCV (hepatitis C virus) non-structural (NS2/3) protease (referred	
CC	to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal	
CC	residue amino acid 810 to 906, or having a minimal amino acid sequence	
CC	from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length	
CC	NS2/3 protease. Also included are (1) a composition (c) comprising an	
CC	isolated HCV NS2/3 protease selected from full length NS2/3 protease, or	
CC	its truncation or a mutated sequence, where the protease is in a solution	
CC	comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)	
CC	to prevent auto-cleavage of the protease, (2) a NS2/3 inhibitory peptide	

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 866-1206
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 341 AA;

Query Match 100.0%; Score 1778; DB 5; Length 341;

Best Local Similarity 100.0%; Pred. No. 9,6e-164;

Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCVHPELFDITKLLAFGLPMTVQAGITVYFVRAQGLIRACMLV 60
 DB 1 VRGRDAIILLTCVHPELFDITKLLAFGLPMTVQAGITVYFVRAQGLIRACMLV 60
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHITPLQDMHAGLDAVAVEPVFSMEVKIIT 120
 DB 61 RKAAGHYVQMAFMKLAALGTYYVDHITPLQDMHAGLDAVAVEPVFSMEVKIIT 120
 QY 121 WGAADTAACGDIISGLPVARSARGREILIGPADNFBGQWRLLAPITAVSQTRGLGCIIT 180
 DB 121 WGAADTAACGDIISGLPVARSARGREILIGPADNFBGQWRLLAPITAVSQTRGLGCIIT 180
 QY 121 WGAADTAACGDIISGLPVARSARGREILIGPADNFBGQWRLLAPITAVSQTRGLGCIIT 180
 DB 121 WGAADTAACGDIISGLPVARSARGREILIGPADNFBGQWRLLAPITAVSQTRGLGCIIT 180
 QY 181 SLTRGDKNOYGEVQVSTATQSPFLATCVCNVCMTVFHAGSKTLAKGKPIQMTYND 240
 DB 181 SLTRGDKNOYGEVQVSTATQSPFLATCVCNVCMTVFHAGSKTLAKGKPIQMTYND 240
 QY 241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTNRADYIPVRRGDSRGLISRPVSYLKSS 300
 DB 241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTNRADYIPVRRGDSRGLISRPVSYLKSS 300
 QY 301 GGPILCPSGHNAVGFRAAIVCTRGVAKAVDFIPVESHMETTYR 341
 DB 301 GGPILCPSGHNAVGFRAAIVCTRGVAKAVDFIPVESHMETTYR 341

RESULT 2

ABG32186 standard; protein; 352 AA.

ABG32186;

05-NOV-2002 (first entry)

HCV protease NS2/3 truncation mutant 855-1206.

XX HCV, enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;
 XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 XX hepatocellular carcinoma; hepatitis C virus; hepatitis C virus;
 XX hepatitis C virus.
 XX Synthetic.

PN WO200248375-A2.
 XX 20-JUN-2002.
 PD 13-DEC-2001; 2001WO-C0A001796.
 XX 15-DEC-2000; 2000US-0256031P.
 PR (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.
 XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;
 PI WPI; 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.

Claim 41; Page 61-62; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 855-1206
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 352 AA;

Query Match 100.0%; Score 1778; DB 5; Length 352;

Best Local Similarity 100.0%; Pred. No. 1e-163;

Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCVHPELFDITKLLAFGLPMTVQAGITVYFVRAQGLIRACMLV 60
 DB 12 VRGRDAIILLTCVHPELFDITKLLAFGLPMTVQAGITVYFVRAQGLIRACMLV 71
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHITPLQDMHAGLDAVAVEPVFSMEVKIIT 120
 DB 72 RKAAGHYVQMAFMKLAALGTYYVDHITPLQDMHAGLDAVAVEPVFSMEVKIIT 131
 QY 121 WGAADTAACGDIISGLPVARSARGREILIGPADNFBGQWRLLAPITAVSQTRGLGCIIT 180
 DB 132 WGAADTAACGDIISGLPVARSARGREILIGPADNFBGQWRLLAPITAVSQTRGLGCIIT 191
 QY 181 SLTRGDKNOYGEVQVSTATQSPFLATCVCNVCMTVFHAGSKTLAKGKPIQMTYND 240

DB 192 SLTRDKNQVEGVSTATQSLATCAGVGCWTFVHGAGSKTLAPKPIQWNTND 251
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGDSRGLSLSPRVSYLKGS 300
 DB 252 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGDSRGLSLSPRVSYLKGS 311
 QY 301 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
 DB 312 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 352

RESULT 3
 ABG32185
 ID ABG32185 standard; protein: 380 AA.
 AC ABG32185;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE HCV protease NS2/3 truncation mutant 827-1206.
 XX
 KM HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 KM hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KM chaotropic agent; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 OS
 PN W0200248375-A2.
 PD 20-JUN-2002.
 XX
 PF 13-DEC-2001; 2001WO-CA001796.
 XX
 PR 15-DEC-2000; 2000US-0256031P.
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;
 DR WPI; 2002-599511/64.
 PT Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.
 XX
 PS Claim 41; Page 60-61; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32185; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an

CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 827-1206
 CC (numbered relative to the full length NS2/3 protein)
 CC
 XX Sequence 380 AA;
 SQ

Query Match 100.0%; Score 1778; DB 5; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.1e-163;
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNGGRDAIILLTCANVPELIPDITKLLAIFGLVWLQAGITKVYFPAQGLIACMLV 60
 DB 40 VNGGRDAIILLTCANVPELIPDITKLLAIFGLVWLQAGITKVYFPAQGLIACMLV 99
 QY 61 RRAAGGHVQMAFMKLAALTGTYYVDHLTPLQDMAHAGLDIAVAVEPIFSDMEVKIT 120
 DB 100 RRAAGGHVQMAFMKLAALTGTYYVDHLTPLQDMAHAGLDIAVAVEPIFSDMEVKIT 159
 QY 121 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGWELLAPITAYSQOTRGLGCIIT 180
 DB 160 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGWELLAPITAYSQOTRGLGCIIT 219
 QY 181 SLTGRDKNQVEGVSTATQSLATCAGVGCWTFVHGAGSKTLAPKPIQWNTND 240
 DB 220 SLTGRDKNQVEGVSTATQSLATCAGVGCWTFVHGAGSKTLAPKPIQWNTND 279
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGDSRGLSLSPRVSYLKGS 300
 DB 280 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGDSRGLSLSPRVSYLKGS 339
 QY 301 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
 DB 340 GGPLCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 380

RESULT 4
 ABG32184
 ID ABG32184 standard; protein: 393 AA.
 AC ABG32184;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 DE HCV protease NS2/3 truncation mutant 815-1206.
 XX
 KM HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 KM hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KM chaotropic agent; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 PN W0200248375-A2.
 PD 20-JUN-2002.
 XX
 PF 13-DEC-2001; 2001WO-CA001796.
 XX
 PR 15-DEC-2000; 2000US-0256031P.
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 PI Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;
 DR WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.

XX Claim 41; Page 59-60; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterization of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The
CC present sequence represents the NS2/3 truncation mutant 815-1206
CC (numbered relative to the full length NS2/3 protein)

XX Sequence 393 AA:

SO Query Match 100.0%; Score 1778; DB 5; Length 393;
Best Local Similarity 100.0%; Pred. No. 1,2e-163;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELIFDTIKMLAIFGPMVLQAGITVPYFVRAQGLIRACMLV 60
DB 53 VRGRDAIILITCAVHPELIFDTIKMLAIFGPMVLQAGITVPYFVRAQGLIRACMLV 112

QY 61 RKAAGHYVQWAFKMLAALGTYYDHLTFLQWMAHGLDLAVNEPVIFSMEXKIT 120
DB 113 RKAAGHYVQWAFKMLAALGTYYDHLTFLQWMAHGLDLAVNEPVIFSMEXKIT 172

QY 121 WGAADTAACGDIISGLPVASARGREIILGPADNFGQGMRLAPITVSOQTRILLCIIT 180
DB 173 WGAADTAACGDIISGLPVASARGREIILGPADNFGQGMRLAPITVSOQTRILLCIIT 232

QY 181 SLTRGRDNOVGEVQVSTATQSHFLATCVNGVCMVFHAGSTLGLPKGPIITQMTTND 240
DB 233 SLTRGRDNOVGEVQVSTATQSHFLATCVNGVCMVFHAGSTLGLPKGPIITQMTTND 292

QY 241 QDLVGMQAPPGASMTCTCGSSDLVYVTHADVPIPRARGDSRGLISRPVSYKGS 300
DB 293 QDLVGMQAPPGASMTCTCGSSDLVYVTHADVPIPRARGDSRGLISRPVSYKGS 352

QY 301 GGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMETTNR 341
DB 353 GGPILCPGSHAVGIFRAAVCTRGVAKAVDPIPVESMETTNR 393

* RESULT 5

ABG32181
ID ABG32181 standard; protein; 409 AA.
XX
XX
AC ABG32181;
XX
XX 05-NOV-2002 (first entry)
XX
XX HCV protease NS2/3 (810-1206).
XX
XX HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
XX chronic liver disease; cirrhosis; end-stage liver disease; virocid;
XX hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
XX chaotropic agent; mutant; mutein.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 398..409
FT /note="Streptavidin tag"
PD WO200248375-A2.
XX 20-JUN-2002.
XX 13-DEC-2001; 2001WO-CA001796.
XX 15-DEC-2000; 2000US-0256031P.
XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX WPI; 2002-599511/64.
XX N-PSDB; ABK90406.
XX
XX Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.

XX Claim 42; Fig 1B; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC residue amino acid 810 to 906, or having a minimal amino acid sequence
CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC NS2/3 protease. Also included are (1) a composition (C) comprising an
CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC its truncation or a mutated sequence, where the protease is in a solution
CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
CC protease, involving isolating the protease in the presence of a
CC chaotropic agent, refolding the isolated protease by contacting it with a
CC reducing agent, and LDAO in the presence of reduced concentration of the
CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC protease, involving diluting refolded inactive NS2/3 protease in a medium
CC containing an activation detergent to induce auto-cleavage of the NS2/3
CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC protease, involving incubating the active NS2/3 protease produced by M2
CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC cleavage products or their fragments, and measuring the presence or
CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC active NS2/3 protease, involving carrying out M3 in the presence of, or
CC absence of the potential inhibitor, comparing the amount of uncleaved
CC NS2/3 protease, cleavage products or their fragments. The protease is
CC useful for detailed biochemical characterization of the enzymes and in
CC the development of in vitro assays for screening novel inhibitors of
CC NS2/3 protease which are useful as therapeutic agents against HCV
CC infection (which causes chronic liver disease, cirrhosis and end-stage
CC liver disease. M1 is useful for high level production of protease. The

CC present sequence represents the NS2/3 (810-1206) protein, which has a C-terminal streptavidin tag

XX Sequence 409 AA;

Query Match 100.0%; Score 1778; DB 5; Length 409;
Best Local Similarity 100.0%; Pred. No. 12e-163;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 60
DB VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 116
QY 61 RKAAGHYVQMAFMKLAALTGTYYVDHLPLODMAHAGLDLAVAVEPVFSDMEVKIIT 120
DB 117 RKAAGHYVQMAFMKLAALTGTYYVDHLPLODMAHAGLDLAVAVEPVFSDMEVKIIT 176
QY 121 WGAADTAACGDIISGLPVSAARRREILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
DB 177 WGAADTAACGDIISGLPVSAARRREILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 236
QY 181 SLTRGRDNQVGEVQVYSTATQSLATCVNGVCTVFGAGSKTLAEPKGPITOMYTNVD 240
DB 237 SLTRGRDNQVGEVQVYSTATQSLATCVNGVCTVFGAGSKTLAEPKGPITOMYTNVD 296
QY 241 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSGSLSPRPVSYLKSS 300
DB 297 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSGSLSPRPVSYLKSS 356
QY 301 GGPILCPSGHAIVGIFRAAVCTRGVAKAVDFIVESMETTR 341
DB 357 GGPILCPSGHAIVGIFRAAVCTRGVAKAVDFIVESMETTR 397

RESULT 6
AAR82694

ID AAR82694 standard; protein; 3010 AA.

XX AAR82694;

AC 16-OCT-2003 (revised)
DT 14-NOV-1996 (first entry)

XX Partial HCV non-structural polyprotein.

DE proteinase; hepatitis C virus; screening; inhibitor; proteolytic;
KW identification; cleavage.

XX Hepatitis C virus; virus.

XX Key Location/Qualifiers
FH 898.1233
FT /note="partial proteinase; see AAR82692"
FT 992.1907
FT /note="partial proteinase; see AAR82693"

XX JP07184648-A.

XX 25-JUL-1995.

XX 05-FEB-1993; 93JP-00018854.

XX 07-FEB-1992; 92JP-00022657.

XX 18-SEP-1992; 92JP-00249240.

XX 04-DEC-1992; 92JP-00325303.

XX (KAEN/) KAENNO K.
PA (SUMO) SUMITOMO METAL IND LTD.
PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.

XX WPI; 1995-287962/38.

XX N-PSDB; AAT03960.

PT An HCV proteinase active substance - which has activity as an anti-HCV agent and can be used to screen for proteinase inhibitors.

XX Disclosure; Page 39-48; 52pp; Japanese.

XX The present sequence is a partial Hepatitis C Virus (HCV) polyprotein from the non-structural region. Partial proteinase sequences (AAR82692-93) are contained within this sequence. The proteinases can be used as anti-HCV agents. They can also be used to screen cdds. for their ability to inhibit their proteolytic activity. In this way proteinase inhibitors can be identified. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 3010 AA;

Query Match 96.2%; Score 1710; DB 2; Length 3010;
Best Local Similarity 94.7%; Pred. No. 8.3e-156;
Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 60
DB 866 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITKVPYFVRAQGLIRACMLV 925
QY 61 RKAAGHYVQMAFMKLAALTGTYYVDHLPLODMAHAGLDLAVAVEPVFSDMEVKIIT 120
DB 926 RKAAGHYVQMAFMKLAALTGTYYVDHLPLODMAHAGLDLAVAVEPVFSDMEVKIIT 985
QY 121 WGAADTAACGDIISGLPVSAARRREILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
DB 986 WGAADTAACGDIISGLPVSAARRREILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 1045
QY 181 SLTRGRDNQVGEVQVYSTATQSLATCVNGVCTVFGAGSKTLAEPKGPITOMYTNVD 240
DB 1046 SLTRGRDNQVGEVQVYSTATQSLATCVNGVCTVFGAGSKTLAEPKGPITOMYTNVD 1105
QY 241 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSGSLSPRPVSYLKSS 300
DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVPYRRRGDSGSLSPRPVSYLKSS 1165
QY 301 GGPILCPSGHAIVGIFRAAVCTRGVAKAVDFIVESMETTR 341
DB 1166 GGPILCPSGHAIVGIFRAAVCTRGVAKAVDFIVESMETTR 1206

RESULT 7
AAR86622

ID AAR86622 standard; protein; 3010 AA.

XX AAR86622;

AC 16-OCT-2003 (revised)
DT 16-OCT-1995 (first entry)

XX HCV protein cleavable with new serine proteinase.

XX proteinase; serine; cleavage; hepatitis C virus; HCV.

XX Hepatitis C virus; virus.

XX Key Location/Qualifiers
FH Cleavage-site 2419..2420
FT /note="serine protease cleavage site"

XX JP06315377-A.

XX 15-NOV-1994.

XX 06-MAY-1993; 93JP-00105666.

XX 06-MAY-1993; 93JP-00105666.

XX (KAEN/) KAENNO K.
PA (SUMO) SUMITOMO METAL IND LTD.
PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.

XX WPI; 1995-032330/05.
 DR N-PSDB; AAQ80498.
 XX
 FT New HCV-originated proteinase active substance - used for site-specific
 FT cleavage by an intermolecular reaction and the purification thereof.
 XX
 PS Disclosure; Page 10-19; 23pp; Japanese.
 XX
 CC This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is
 CC cleaved between amino acids 2419 and 2420, by a new serine protease,
 CC contg. the sequence of AAR68621. The proteinase is purified as a fused
 CC product with the dihydrofolate reductase protein by using a methotrexate
 CC column. It can be used for the development of an inhibitor for HCV
 CC proteinase. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 SO Sequence 3010 AA;
 Query Match 95.8%; Score 1705; DB 2; Length 3010;
 Best Local Similarity 94.4%; Pred. No. 2.5e-155;
 Matches 322; Conservative 11; Mismatches 8; Indels 0; Gaps 0;
 QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLOAGITRVYFVRAQGLIRACMLV 60
 DB 866 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLOAGITRVYFVRAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLPLODVAHAGLRDLAVAVEPVFSDMEVKIIT 120
 DB 926 RKAAGHYVQMAFMKLAALGTYYVDHLPLODVAHAGLRDLAVAVEPVFSDMEVKIIT 985
 QY 121 WGAADTAACGDIISGLPVSARRGRRIILGPADNFGQGRRLIAPITAVSOOTRGLIGCIIT 180
 DB 986 WGAADTAACGDIISGLPVSARRGRRIILGPADNFGQGRRLIAPITAVSOOTRGLIGCIIT 1045
 QY 181 SLTRGRDNQYGEVQVNSTAQSFLATCVNGVCTVPHAGSKITLAGPKGPIITOMYTNDV 240
 DB 1046 SLTRGRDNQYGEVQVNSTAQSFLATCVNGVCTVPHAGSKITLAGPKGPIITOMYTNDV 1105
 QY 241 ODVGMQAPRGARSMPTCTCGSSDLVLTTHADVIPRRGRSDRGSLISRPVSYLKGS 300
 DB 1106 QDLVGMWAPRGARSMPTCTCGSSDLVLTTHADVIPRRGRSDRGSLISRPVSYLKGS 1165
 QY 301 GGPIILCPSGHAGVIFRAAVCTRGVAKAVDEIPVSMETTNR 341
 DB 1166 GGPIILCPSGHAGVIFRAAVCTRGVAKAVDEIPVSMETTNR 1206
 RESULT 8
 AAR6864
 ID AAR6864 standard; protein; 3010 AA.
 XX
 AC AAR6864;
 XX
 DT 06-DEC-1995 (first entry)
 XX
 DE Hepatitis C virus RNA helicase.
 XX
 KW Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;
 KW baculovirus; recombinant production.
 XX
 OS Hepatitis C virus.
 XX
 FH Key
 FT 196. 198
 FT /label= N-linked glycosylation site
 FT 209. 211
 FT /label= N-linked glycosylation site
 FT 234. 236
 FT /label= N-linked glycosylation site
 FT 250. 252
 FT /label= N-linked glycosylation site
 FT 305. 307
 FT /label= N-linked glycosylation site

FT Region 325. 327
 FT /label= N-linked glycosylation site
 FT Region 417. 419
 FT /label= N-linked glycosylation site
 FT Region 423. 425
 FT /label= N-linked glycosylation site
 FT Region 430. 432
 FT /label= N-linked glycosylation site
 FT Region 448. 450
 FT /label= N-linked glycosylation site
 FT Region 532. 534
 FT /label= N-linked glycosylation site
 FT Region 556. 558
 FT /label= N-linked glycosylation site
 FT Region 576. 578
 FT /label= N-linked glycosylation site
 FT Region 623. 625
 FT /label= N-linked glycosylation site
 FT Region 645. 647
 FT /label= N-linked glycosylation site
 FT Region 1213. 1215
 FT /label= N-linked glycosylation site
 FT Region 1255. 1257
 FT /label= N-linked glycosylation site
 FT Region 2041. 2043
 FT /label= N-linked glycosylation site
 FT Region 2077. 2079
 FT /label= N-linked glycosylation site
 FT Region 2240. 2242
 FT /label= N-linked glycosylation site
 FT Region 2788. 2790
 FT /label= N-linked glycosylation site
 PN JP06319583-A.
 XX
 PD 22-NOV-1994.
 XX
 PF 18-SEP-1992; 92JP-00249241.
 XX
 PR 18-SEP-1992; 92JP-00249241.
 XX
 PA (SOYA-) SOYAKU GIJUTSU KENKUSHO KK.
 XX
 DR WPI; 1995-040330/06.
 DR N-PSDB; AAQ81559.
 XX
 PT of hepatitis C virus helicase gene in baculovirus - useful for large
 PT scale prodn. of RNA helicase.
 PS Claim 1; Fig 1-4; 9pp; Japanese.
 PS
 CC AAQ81559 encodes AAR6864 hepatitis C virus (HCV) RNA helicase. The DNA
 CC was used in the construction of an expression vector, which was used to
 CC transform a baculovirus host. The transformed baculovirus could then be
 CC used for the recombinant prodn. of HCV RNA helicase
 XX
 SQ Sequence 3010 AA;
 Query Match 95.8%; Score 1704; DB 2; Length 3010;
 Best Local Similarity 94.4%; Pred. No. 3.2e-155;
 Matches 322; Conservative 11; Mismatches 8; Indels 0; Gaps 0;
 QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLOAGITRVYFVRAQGLIRACMLV 60
 DB 866 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVLOAGITRVYFVRAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLPLODVAHAGLRDLAVAVEPVFSDMEVKIIT 120
 DB 926 RKAAGHYVQMAFMKLAALGTYYVDHLPLODVAHAGLRDLAVAVEPVFSDMEVKIIT 985
 QY 121 WGAADTAACGDIISGLPVSARRGRRIILGPADNFGQGRRLIAPITAVSOOTRGLIGCIIT 180
 DB 986 WGAADTAACGDIISGLPVSARRGRRIILGPADNFGQGRRLIAPITAVSOOTRGLIGCIIT 1045

QY 181 SLTGRDNQVGEVGVSTATQSFATCVNGVCMVTFHGAAGSKTLAAGPKPITQMTNTVD 240
 DB 1046 SLTGRDNQVGEVGVSTATQSFATCVNGVCMVTFHGAAGSKTLAAGPKPITQMTNTVD 1105
 QY 241 QDLVGMQAPPGARSMTPTCTCGSSDLYLVTRHADVPVRRRGRSGSLSPRPVSYLKSS 300
 DB 1106 QDLVGMQAPPGARSMTPTCTCGSSDLYLVTRHADVPVRRRGRSGSLSPRPVSYLKSS 1165
 QY 301 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
 DB 1166 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206
 RESULT 9
 ABG30601 standard; protein; 2201 AA.
 AC ABG30601;
 XX
 DT 21-OCT-2002 (first entry)
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 882 /label= Arg, Lys
 FT Misc-difference 2183 /note= "Wild type Met substituted by Thr"
 FT
 XX WO200252015-A2.
 PD 04-JUL-2002.
 PF 20-DEC-2001; 2001WO-CA001843.
 PR 22-DEC-2000; 2000US-0257857P.
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukulj G, Pause A;
 DR WPI; 2002-575382/61.
 XX
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 XX Claim 3; Page; 140pp; English.
 XX
 CC The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis
 CC C virus replicon Apk12 and contains the viral protease NS2/3, protease
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
 CC This sequence does not appear in the specification but has been created
 CC from the wild type sequence shown in ABG30580 using information given in

CC the claims of the invention
 XX SQ Sequence 2201 AA;
 Query Match 95.7%; Score 1702; DB 5; Length 2201;
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
 QY 1 VAGSDAIIILTCANHPHLEIPITKLLAIIFGLMVAQAGTKVYFPRAGLRACMLV 60
 DB 57 VAGSDAIIILTCANHPHLEIPITKLLAIIFGLMVAQAGTKVYFPRAGLRACMLV 116
 QY 61 RRAAGHYVQMAFMKLAALGTGYVDHPLPLQDMNAGLRDLAVAVEVIESDMEVKIT 120
 DB 117 RYVAGHYVQMAFMKLAALGTGYVDHPLPLQDMNAGLRDLAVAVEVIESDMEVKIT 176
 QY 121 WGADTPAACGDIISGLPVARRRGEIILGPADNFBEGGWRLLAPITAYGQOTRGLGCIIT 180
 DB 177 WGADTPAACGDIISGLPVARRRGEIILGPADNFBEGGWRLLAPITAYGQOTRGLGCIIT 236
 QY 181 SLTGRDNQVGEVGVSTATQSFATCVNGVCMVTFHGAAGSKTLAAGPKPITQMTNTVD 240
 DB 237 SLTGRDNQVGEVGVSTATQSFATCVNGVCMVTFHGAAGSKTLAAGPKPITQMTNTVD 296
 QY 241 QDLVGMQAPPGARSMTPTCTCGSSDLYLVTRHADVPVRRRGRSGSLSPRPVSYLKSS 300
 DB 297 QDLVGMQAPPGARSMTPTCTCGSSDLYLVTRHADVPVRRRGRSGSLSPRPVSYLKSS 356
 QY 301 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
 DB 357 GGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 397
 RESULT 10
 ABG30591 standard; protein; 2201 AA.
 ID ABG30591
 XX
 AC ABG30591;
 XX
 DT 21-OCT-2002 (first entry)
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 751 /note= "Wild type Ser substituted by Gly"
 FT Misc-difference 882 /label= Arg, Lys
 FT
 XX WO200252015-A2.
 PD 04-JUL-2002.
 PF 20-DEC-2001; 2001WO-CA001843.
 PR 22-DEC-2000; 2000US-0257857P.
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukulj G, Pause A;
 DR WPI; 2002-575382/61.
 XX
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.

PS Claim 3; Page; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-
 CC replicating hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis
 CC C virus replicon APOK12 and contains the viral protease NS2/3, protease
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
 CC This sequence does not appear in the specification but has been created
 CC from the wild type sequence shown in ABG30580 using information given in
 CC the claims of the invention

XX Sequence 2201 AA:

Query Match 95.7%; Score 1702; DB 5; Length 2201;

Best Local Similarity 94.1%; Pred. No. 3.2e-155; Mismatches 9; Indels 0; Gaps 0;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIILTCVHPELFDITKLLAFGLPMLVQAGITKVPYFVRAQGLIRACMLV 60
 DB 57 VRGGRDAVILITCAIHPELFTITKLLALIGPLMVQAGITKVPYFVRAHGLIRACMLV 116
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDMEVKIIT 120
 DB 117 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDMEVKIIT 176
 QY 121 WGAADTAACGDIISGLPVSAARGREIILGPADNFGGQWRLLAPITAYSOQTRGLGCIIT 180
 DB 177 WGAADTAACGDIISGLPVSAARGREIILGPADNFGGQWRLLAPITAYSOQTRGLGCIIT 236
 QY 181 SLTGRDNQVGEVQVSTATQSFATCVGVCMTVFHGAQSKTLAPKGPITOMYTND 240
 DB 237 SLTGRDNQVGEVQVSTATQSFATCVGVCMTVFHGAQSKTLAPKGPITOMYTND 296
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 300
 DB 297 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 356
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPFVPSMETTKR 341
 DB 357 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPFVPSMETTKR 397

RESULT 11

ABG30600 standard; protein; 2201 AA.

XX ABG30600;

XX 21-OCT-2002 (first entry)

XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

XX cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; protein.

OS Hepatitis C virus.

XX Synthetic.

XX Key Location/Qualifiers

FT Msc-difference 882

FT Msc-difference 1357

FT /note= "Wild type Pro substituted by Leu"

XX WO200252015-A2.

XX 04-JUL-2002.

XX 20-DEC-2001; 2001WO-CN001843.

XX 22-DEC-2000; 2000US-0257857P.

XX (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.

XX Kukulj G, Pause A;

XX WPI, 2002-575382/61.

XX New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.

PS Claim 3; Page; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide, and a 3'-NTR region. The self-
 CC replicating hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis
 CC C virus replicon APOK12 and contains the viral protease NS2/3, protease
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
 CC This sequence does not appear in the specification but has been created
 CC from the wild type sequence shown in ABG30580 using information given in
 CC the claims of the invention

XX Sequence 2201 AA:

Query Match 95.7%; Score 1702; DB 5; Length 2201;

Best Local Similarity 94.1%; Pred. No. 3.2e-155; Mismatches 9; Indels 0; Gaps 0;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIILTCVHPELFDITKLLAFGLPMLVQAGITKVPYFVRAQGLIRACMLV 60
 DB 57 VRGGRDAVILITCAIHPELFTITKLLALIGPLMVQAGITKVPYFVRAHGLIRACMLV 116
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDMEVKIIT 120
 DB 117 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPVFSDMEVKIIT 176
 QY 121 WGAADTAACGDIISGLPVSAARGREIILGPADNFGGQWRLLAPITAYSOQTRGLGCIIT 180
 DB 177 WGAADTAACGDIISGLPVSAARGREIILGPADNFGGQWRLLAPITAYSOQTRGLGCIIT 236
 QY 181 SLTGRDNQVGEVQVSTATQSFATCVGVCMTVFHGAQSKTLAPKGPITOMYTND 240
 DB 237 SLTGRDNQVGEVQVSTATQSFATCVGVCMTVFHGAQSKTLAPKGPITOMYTND 296
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 300
 DB 297 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIRRRGGDSGSLSPRPVSYLKSS 356
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPFVPSMETTKR 341
 DB 357 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPFVPSMETTKR 397

RESULT 12

ABG30581

ID ABG30581 standard; protein; 2201 AA.
 XX
 AC ABG30581;
 XX
 DT 21-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.
 XX
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B.
 XX
 OS Hepatitis C virus.
 XX
 PN W0200252015-A2.
 XX
 PD 04-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-CA001843.
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukulj G, Pause A;
 XX
 DR WPI; 2002-575382/61.
 DR N-PSDB; ABK88573.
 XX
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 PS Disclosure; Page 49-58; 140pp; English.
 XX
 XX The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
 CC replicon ApGx12 and contains the viral protease NS2/3, protease complex
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
 CC
 XX
 SQ Sequence 2201 AA;
 XX
 Query Match 95.7%; Score 1702; DB 5; Length 2201;
 Best Local Similarity 94.1%; Pred. No. 3.2e-155; Indels 0; Gaps 0;
 Matches 321; Conservative 11; Mismatches 9;
 QY 1 VRGGRDAIILITCAVHELEIFDITKLLAIFGPIMLVLAQITKVPYFVRAGLIRACMLV 60
 DB 57 VRGGRDAVILITCAIHELEIFDITKLLAIFGPIMLVLAQITKVPYFVRAGLIRACMLV 116
 QY 61 RKAAGHYVQMAEFKALATGTYVDLTPQDMANHGDLAVAVPVEFMEVYIIT 120
 DB 117 RKAAGHYVQMAEFKALATGTYVDLTPQDMANHGDLAVAVPVEFMEVYIIT 176
 QY 121 WGAATAACGDIISGLPYASARGREIILGPAIDNEFGQGRLLAPITAYSOOTRGLGCIIT 180
 DB 177 WGAATAACGDIISGLPYASARGREIILGPAIDNEFGQGRLLAPITAYSOOTRGLGCIIT 236
 QY 181 SLTGRDKNOVGEVQVSTATQSLATCVNGVCTVHGAGSKTLAPKPIQMYNNVD 240
 DB 237 SLTGRDKNOVGEVQVSTATQSLATCVNGVCTVHGAGSKTLAPKPIQMYNNVD 296
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVI PVRRRGRSGSLSPRPVSYLKSS 300

DB 297 QDLVGMQAPPGARSMPTCTCGSSDLYLVTRHADVI PVRRRGRSGSLSPRPVSYLKSS 356
 QY 301 GGFLPCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341
 DB 357 GGFLPCPSGHAAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 397
 RESULT 13
 ID ABG30593
 XX ABG30593 standard; protein; 2201 AA.
 AC ABG30593;
 XX
 DT 21-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #4.
 XX
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.
 XX
 OS Hepatitis C virus.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 882 /label= Arg, Lys
 FT Misc-difference 892 /note= "Wild type Leu substituted by Phe"
 FT
 XX
 PN W0200252015-A2.
 XX
 PD 04-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-CA001843.
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEHR) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukulj G, Pause A;
 XX
 DR WPI; 2002-575382/61.
 XX
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 PS Claim 3; Page; 140pp; English.
 XX
 XX The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)/C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis
 CC C virus replicon ApGx12 and contains the viral protease NS2/3, protease
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:
 CC This sequence does not appear in the specification but has been created
 CC from the wild type sequence shown in ABG30580 using information given in
 CC the claims of the invention
 CC
 XX
 SQ Sequence 2201 AA;
 XX
 Query Match 95.7%; Score 1702; DB 5; Length 2201;
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCNAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 60
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this
 CC sequence has been created from replicon ApGK12 shown in ABG30581

Db 57 VRGRDAVILLTCNAHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 116
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGLRDLAVAEVYFSDMEVKIIT 120
 Db 117 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGLRDLAVAEVYFSDMEVKIIT 176
 QY 121 WGAADTAACGDIISGLPVSARRREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 180
 Db 177 WGAADTAACGDIISGLPVSARRREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 236
 QY 181 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAAGPKPITQMTNVD 240
 Db 237 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAAGPKPITQMTNVD 296
 QY 241 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRDSGSLSPRPVSYLKSS 300
 Db 297 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRDSGSLSPRPVSYLKSS 356
 QY 301 GGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 341
 Db 357 GGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 397

RESULT 14
 ABG30582
 ID ABG30582 standard; protein; 2201 AA.

AC ABG30582;
 XX
 DT 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.
 XX
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; muten.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 882
 FT /note= "Wild type Lys substituted by Lys or Arg"
 FT Misc-difference 1233
 FT /note= "Wild type Gly substituted by Cys"
 XX
 FN W0200252015-A2.
 XX
 PD 04-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-CA001843.
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PI Kukulj G, Pause A;
 XX
 DR WPI: 2002-575382/61.
 DR N-PSDB; ABK88574.
 XX
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
 PT possess enhanced transduction or replication efficiency, useful for
 PT evaluating potential inhibitors of HCV replication.
 XX
 PS Disclosure; Page 59-69; 140pp; English.

XX The invention describes a self-replicating hepatitis C virus (HCV)
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),
 CC where guanine at position 1 is substituted for adenine, a HCV polypeptide
 CC region coding for a HCV polypeptide; and a 3'-NTR region. The self-

CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also
 CC useful for efficiently establishing cell culture replication. The self-
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at
 CC position 1 is substituted for A, and therefore provides an alternative to
 CC existing systems comprising a self-replicating HCV RNA molecule that, in
 CC conjunction with mutations in the HCV non-structural region, such as the
 CC G(2042)C/R mutations, transduces and/or replicates with greater
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
 CC replicon ApGK12 and contains the viral protease NS2/3, protease complex
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this
 CC sequence has been created from replicon ApGK12 shown in ABG30581

SO Sequence 2201 AA:

Query Match 95.7%; Score 1702; DB 5; Length 2201;
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCNAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 60
 Db 57 VRGRDAVILLTCNAHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 116
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGLRDLAVAEVYFSDMEVKIIT 120
 Db 117 RKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGLRDLAVAEVYFSDMEVKIIT 176
 QY 121 WGAADTAACGDIISGLPVSARRREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 180
 Db 177 WGAADTAACGDIISGLPVSARRREIILGPADNFGOGWRLIAPITAYSOQTRGLGCIIT 236
 QY 181 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAAGPKPITQMTNVD 240
 Db 237 SLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAAGPKPITQMTNVD 296
 QY 241 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRDSGSLSPRPVSYLKSS 300
 Db 297 ODVGMQAPPGARSMETCTCGSSDLYLVTRHADVIPIRRRDSGSLSPRPVSYLKSS 356
 QY 301 GGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 341
 Db 357 GGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTKR 397

RESULT 15
 ABG30580
 ID ABG30580 standard; protein; 2201 AA.

AC ABG30580;
 XX
 DT 21-OCT-2002 (first entry)

DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.
 XX
 KM Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;
 KM cell culture replication; NS2/3; NS3/4; NS3; NS5B.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 882
 FT /note= "Encoded by ARG"
 FT
 FN W0200252015-A2.
 XX
 PD 04-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-CA001843.
 XX
 PR 22-DEC-2000; 2000US-0257857P.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX

PI Kukolj G, Pause A;
XX
DR WPI; 2002-575382/61.
XX
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which
PT possess enhanced transduction or replication efficiency, useful for
PT evaluating potential inhibitors of HCV replication.
XX
PS Disclosure; Page 69-74; 140pp; English.
XX
CC The invention describes a self-replicating hepatitis C virus (HCV)
CC polynucleotide molecule comprising a 5'-non translated region (NTR),
CC where guanine at position 1 is substituted for adenine, a HCV polyprotein
CC region coding for a HCV polyprotein; and a 3'-NTR region. The self-
CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating
CC potential inhibitors of HCV replication. The HCV RNA molecule is also
CC useful for efficiently establishing cell culture replication. The self-
CC replicating polynucleotide molecule contains a 5'-NTR, where G at
CC position 1 is substituted for A, and therefore provides an alternative to
CC existing systems comprising a self-replicating HCV RNA molecule that, in
CC conjunction with mutations in the HCV non-structural region, such as the
CC G(2042)/C/R mutations, transduces and/or replicates with greater
CC efficiency. This amino acid sequence is encoded by the hepatitis C virus
CC replicon ApGK12 and contains the viral protease NS2/3, protease complex
CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B
XX
SQ Sequence 2201 AA;

Query Match 95.7%; Score 1702; DB 5; Length 2201;
Best Local Similarity 94.1%; Pred. No. 3,2e-155;
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIILLTCVAPHELLFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 60
DB 57 VRGGRDAVILLTCVAPHELLFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 116
QY 61 RKVAGHYVQMAKMLALGTIVYDHTPLQDWAHAGLRDLAVAVEPVI FSDMEVKIIT 120
DB 117 RKVAGHYVQMAKMLALGTIVYDHTPLQDWAHAGLRDLAVAVEPVI FSDMEVKIIT 176
QY 121 WGAADTAACGDIISGLPVASARGREILGPADNPEGQGMRLAPITAYSOQTRGLGCIIT 180
DB 177 WGAADTAACGDIISGLPVASARGREILGPADNPEGQGMRLAPITAYSOQTRGLGCIIT 236
QY 181 SLTGRDKNOVEGEVQVYSTATQSLATCNGVCWTFVHAGSKTLAGPKPITQMTNVND 240
DB 237 SLTGRDKNOVEGEVQVYSTATQSLATCNGVCWTFVHAGSKTLAGPKPITQMTNVND 296
QY 241 QDLVGMQAPPGARSMTPTCCSSDLVYVTRHADVI PVRRGDSRGSLLSPRVSYLKSS 300
DB 297 QDLVGMQAPPGARSMTPTCCSSDLVYVTRHADVI PVRRGDSRGSLLSPRVSYLKSS 356
QY 301 GGPILCPSGHAGVIFRAVCTRGVAKAVDPI PVESMETTMR 341
DB 357 GGPILCPSGHAGVIFRAVCTRGVAKAVDPI PVESMETTMR 397

Search completed: May 6, 2004, 09:30:48
Job time : 47.4159 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 11.1056 Seconds

(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-14

Sequence: 1 VRGGRDAIILITCAVHEPLT.....RGVAKAVDFIPVSEMETMKR 341

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1711	96.2	3010	1 A45573	genome polyprotein
2	1710	96.2	3010	1 GNMWVJ	genome polyprotein
3	1685	94.8	3010	1 GNMWTV	genome polyprotein
4	1665	93.6	3010	1 S18030	genome polyprotein
5	1649	92.7	3010	1 GNMWTC	genome polyprotein
6	1565	88.0	3011	1 GNMWV3	genome polyprotein
7	1560	87.7	3011	1 S40770	genome polyprotein
8	1555	87.5	3011	1 GNMWVH	genome polyprotein
9	1362	76.6	3014	1 JC5620	genome polyprotein
10	1269	71.4	3033	1 J01303	genome polyprotein
11	1262	71.0	3033	1 GNMWV8	genome polyprotein
12	408.5	23.0	3005	2 T08841	polyprotein - dou
13	342.5	19.3	2970	2 T08839	polyprotein - marm
14	102.5	5.8	660	2 VHMW32	structural protein
15	101	5.7	600	2 B46642	signal recognition
16	101	5.7	600	2 B46642	DNA-directed DNA p
17	100.5	5.7	353	2 G87392	conserved hypochet
18	99.5	5.6	692	2 H71426	hypothetical prote
19	98	5.5	399	2 AH3038	conserved hypochet
20	97.5	5.5	399	2 G98247	hypothetical 50.8k
21	95.5	5.4	1085	2 T03531	cobn protein homol
22	95.5	5.4	470	2 JC4098	tetracycline 6-hyd
23	93.5	5.3	1380	2 T18309	receptor-adenylate
24	93	5.2	7463	2 T36248	CDA peptide synthe
25	92.5	5.2	706	2 S33761	transferrin precur
26	92.5	5.2	716	2 G83612	hypothetical prote
27	92	5.2	659	2 B44212	structural protein
28	91.5	5.1	3414	1 GNMWVE	genome polyprotein
29	91	5.1	904	2 A84212	hypothetical prote

30	90.5	5.1	668	2 H81775	acconitate hydratase
31	90	5.1	2796	2 JC4743	fatty-acid synthase
32	89.5	5.0	961	2 AE0375	probable cation-tr
33	88.5	5.0	652	2 C70688	probable nitrate r
34	87.5	4.9	3069	2 H70656	fatty-acid synthase
35	87.5	4.9	3412	1 GNMWTV	genome polyprotein
36	86.5	4.9	347	2 S4167	maleate dehydrogena
37	86.5	4.9	470	1 NMTVW8	exo-alpha-sialidas
38	86.5	4.9	707	2 D84154	cadmium-transporti
39	86	4.8	338	1 S34984	UDPglucose 4-epime
40	85.5	4.8	348	2 H70549	probable pdhs prot
41	85.5	4.8	3076	2 A87058	fatty acid synthase
42	85	4.8	470	1 NMTV9	exo-alpha-sialidas
43	84.5	4.8	1057	2 T38171	hugt protein - hum
44	84	4.7	233	2 T35594	hypothetical prote
45	84	4.7	423	2 S43967	p54-alpha stress-a

ALIGNMENTS

RESULT 1

A45573 genome polyprotein - hepatitis C virus (strain JF)

N/Contains: capsid protein C; envelope protein M; hepacivirin (BC 3.4.21.98) (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C/Species: hepatitis C virus

C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

R/Accession: A45573

R/Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: A

Virus Res. 23, 39-53, 1992

A/Reference number: A45573; PMID:92295714; PMID:1318627

A/Accession: A45573

A/Status: preliminary

A/Molecule type: DNA

A/Residue: 1-3010 <TAN>

A/Cross-references: GB:DI1168; GB:DD1171; NID:G221612; PIDN:BAA01943.1; PID:G221613

A/Experimental source: HCV-JT

A/Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:106207)

C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polyprotein; serine

F/2-115/Product: capsid protein C #status predicted <CPC>

F/116-191/Product: envelope protein M #status predicted <EMP>

F/192-389/Product: major envelope protein E #status predicted <NE1>

F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>

F/1230-1237/Region: nucleotide-binding motif A (P-loop)

F/1312-1317/Region: nucleotide-binding motif B

F/1316-1319/Region: DEXH motif

F/1616-1662/Product: nonstructural protein NS4 #status predicted <NS4>

F/1663-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 96.2%; Score 1711; DB 1; Length 3010;

Best local similarity 95.3%; Pred. No. 4.7e-136;

Matches 325; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY	1	VGGGRDAIILITCAVHEPLTITLLAIFRPLVNLQAGTKVYFPAQSLIRACMLV	60
DB	866	VGGGRDAIILITCAVHEPLTITLLAIFRPLVNLQAGTKVYFPAQSLIRACMLV	925
QY	61	RRAGGHVQVAFKMLAALITGVVYDHTPLQDMNAAGLDLAFAVEVIFSDMEVKIT	120
DB	926	RRAGGHVQVAFKMLAALITGVVYDHTPLQDMNAAGLDLAFAVEVIFSDMEVKIT	985
QY	121	WADDTAAAGDITISGIPVSAARGREITLGPADNFEQGRMLAIPITAYGQCRGLGCTT	180
DB	986	WADDTAAAGDITISGIPVSAARGREITLGPADNFEQGRMLAIPITAYGQCRGLGCTT	1045
QY	181	SLTGRDKVQEGEVVSTATOSPLATCVNGCMTVFAGAGSKTLGKPGITQVYTVND	240

Db 1046 SLTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAAGSKTLAAGPKPITOMKTNVD 1105
 QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYTFHADVIPIVRRGDSRGLSPRPVSYLKSS 300
 Db 1106 QDLVGMQAPPGARSMTPTCGSSDLVYTFHADVIPIVRRGDSRGLSPRPVSYLKSS 1165
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
 Db 1166 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 2

GNMWCT
 genome polypeptide - hepatitis C virus (strain J)
 N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
 C:Accession: A39253; PS0086
 R:Kato, N.; Hijioka, M.; Ootsuyama, Y.; Nakagawa, M.; Okoshi, S.; Sugimura, T.; Shimoto, Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
 A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
 A:Reference number: A39253; MUID:91088550; PMID:2175903
 A:Accession: A39253
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <KAT>
 A:Cross-references: GB:090208; NID:9221610; PIDN:BA1423.1; PID:9221611
 R:Kato, N.; Okoshi, S.; Shimotohno, K.
 Proc. Jpn. Acad. 65B, 219-223, 1989
 A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variability
 A:Reference number: PS0085
 A:Accession: PS0086
 A:Molecule type: genomic RNA
 A:Residues: 2650-2707 <KAT>
 A:Experimental source: Japanese isolate
 C:Comment: The cleavage sites of this polypeptide have not been determined.
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
 F:115-191/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <ME>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Product: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 96.2%; Score 1710; DB 1; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 5.7e-136;
 Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLALIFGLMVLQAGITKVPYFRAQGLIRACMLV 60
 Db 866 VRGRDAIILITCAVHPELFDITKLALIFGLMVLQAGITKVPYFRAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAFMKLAALITGYVDHLTPLODMAHAGLRLAAYEVPVFSDEMTKIT 120
 Db 926 RKAAGHYVQMAFMKLAALITGYVDHLTPLODMAHAGLRLAAYEVPVFSDEMTKIT 985
 QY 121 WGAATTAACGDIISGLPVSARGREILIGPADNFEQGRMLLAPITAYSOOTRGLGCIIT 180
 Db 986 WGAATTAACGDIISGLPVSARGREILIGPADNFEQGRMLLAPITAYSOOTRGLGCIIT 1045
 QY 181 SLTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAAGSKTLAAGPKPITOMKTNVD 240
 Db 1046 SLTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAAGSKTLAAGPKPITOMKTNVD 1105
 QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYTFHADVIPIVRRGDSRGLSPRPVSYLKSS 300

Db 1106 QDLVGMQAPPGARSMTPTCGSSDLVYTFHADVIPIVRRGDSRGLSPRPVSYLKSS 1165
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
 Db 1166 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 3

GNMWCT
 genome polypeptide - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (HCV) 3.4.21.98 (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40244
 R:Chen, P.-J.; Lin, M.-H.; Tai, K.-F.; Liu, P.-C.; Lin, C.-J.; Chen, D.-S.
 Virology 188, 102-113, 1992
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449
 A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 A:Cross-references: GB:N64754
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:115-191/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <ME>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 94.8%; Score 1685; DB 1; Length 3010;
 Best Local Similarity 92.7%; Pred. No. 7.5e-134;
 Matches 316; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLALIFGLMVLQAGITKVPYFRAQGLIRACMLV 60
 Db 866 VRGRDAIILITCAVHPELFDITKLALIFGLMVLQAGITKVPYFRAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAFMKLAALITGYVDHLTPLODMAHAGLRLAAYEVPVFSDEMTKIT 120
 Db 926 RKAAGHYVQMAFMKLAALITGYVDHLTPLODMAHAGLRLAAYEVPVFSDEMTKIT 985
 QY 121 WGAATTAACGDIISGLPVSARGREILIGPADNFEQGRMLLAPITAYSOOTRGLGCIIT 180
 Db 986 WGAATTAACGDIISGLPVSARGREILIGPADNFEQGRMLLAPITAYSOOTRGLGCIIT 1045
 QY 181 SLTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAAGSKTLAAGPKPITOMKTNVD 240
 Db 1046 SLTGRDKNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAAGSKTLAAGPKPITOMKTNVD 1105
 QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYTFHADVIPIVRRGDSRGLSPRPVSYLKSS 300
 Db 1106 QDLVGMQAPPGARSMTPTCGSSDLVYTFHADVIPIVRRGDSRGLSPRPVSYLKSS 1165
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
 Db 1166 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206
 RESULT 4
 S18030
 genome polypeptide - hepatitis C virus (isolate JKI)

N;Contains: capsid protein C, envelope protein M, hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
 C;Species: hepatitis C virus
 A;Variety: isolate UK1
 C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
 C;Accession: S18030; S33570; A48332; S18029
 R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 Submitted to the EMBL Data Library, September 1991
 A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient
 A;Reference number: S18028
 A;Accession: S18030
 A;Molecule type: genomic RNA
 A;Residues: 1-3010 <HON>
 A;Cross-references: EMBL:X61596; NID:959478; PIDN:CAA3793.1; PID:559479
 A;Experimental source: isolate UK1 from an individual
 R;Honda, M.; Kaneko, S.; Uoura, M.; Kobayashi, K.; Murakami, S.
 Arch. Virol. 128, 163-169, 1993
 A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
 A;Reference number: A48332; MUID:93119270; PMID:8380322
 A;Accession: S33570
 A;Molecule type: genomic RNA
 A;Residues: 1-547; 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HON>
 A;Cross-references: EMBL:X61591
 A;Note: this sequence is inconsistent with the nucleotide translation
 A;Note: the authors translated the codon AGG for residue 43 as Pro. TCG for residue 320 as Trp, and TTC for residue 771 as Ser
 A;Note: Sequence extracted from NCBI database (NCBI:121747, NCBI:P.121748)
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F;116-191/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: capsid protein C #status predicted <EPM>
 F;116-191/Product: major envelope protein M #status predicted <MEP>
 F;116-191/Product: major envelope protein E #status predicted <NEE>
 F;330-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1635/Product: nonstructural protein NS3 #status predicted <NS3>
 F;1230-1237/Product: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: DEXH motif B
 F;1316-1319/Region: DEXH motif B
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 93.6%; Score 1665; DB 1; Length 3010;
 Best Local Similarity 92.7%; Pred. No. 3.7e-132;
 Matches 316; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

1 VGGRAIILITCAVHEPILFDITKLLAFGLMVLQAGITVFPVFAAGIIRACMY 60
 866 VGGRAIILITCAVHEPILFDITKLLAFGLMVLQAGITVFPVFAAGIIRACMY 925
 61 RKAAGHYVQMAFMKLAALGTGYVDHLPLQDMAHAGLDLVAVEPIFSMEVKIIT 120
 926 RKAAGHYVQMAFMKLAALGTGYVDHLPLQDMAHAGLDLVAVEPIFSMEVKIIT 985
 121 WGAADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOOTRGLGCIIT 180
 986 WGAADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOOTRGLGCIIT 1045
 181 SLTGRDKQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 240
 1046 SLTGRDKQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 1105
 241 QDIVGMQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 300
 1106 QDIVGMQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 1165
 301 GGLLCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTR 341
 1166 GGLLCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTR 1206

RESULT 5

GNMVC

genome polyprotein - hepatitis C virus
 N;Contains: capsid protein C, envelope protein M, hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
 C;Species: hepatitis C virus
 C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 A;Accession: A38465
 R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991
 A;Title: Structure and organization of the hepatitis C virus genome isolated from human
 A;Reference number: A38465; MUID:91140698; PMID:1847440
 A;Accession: A38465
 A;Molecule type: genomic RNA
 A;Residues: 1-3010 <TAK>
 A;Cross-references: EMBL:M58335; NID:9329770; PIDN:AA72945.1; PID:9329771
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; capsid protein C, envelope protein M; glycoprotein; hydrolase; nonstructural
 F;116-191/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: capsid protein C #status predicted <EPM>
 F;116-191/Product: major envelope protein M #status predicted <MEP>
 F;116-191/Product: major envelope protein E #status predicted <NEE>
 F;330-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1635/Product: nonstructural protein NS3 #status predicted <NS3>
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif B
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 92.7%; Score 1649; DB 1; Length 3010;
 Best Local Similarity 92.1%; Pred. No. 8.3e-131;
 Matches 313; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

2 RGGRAIILITCAVHEPILFDITKLLAFGLMVLQAGITVFPVFAAGIIRACMY 61
 867 RGGRAIILITCAVHEPILFDITKLLAFGLMVLQAGITVFPVFAAGIIRACMY 926
 62 KKAAGHYVQMAFMKLAALGTGYVDHLPLQDMAHAGLDLVAVEPIFSMEVKIIT 121
 927 KKAAGHYVQMAFMKLAALGTGYVDHLPLQDMAHAGLDLVAVEPIFSMEVKIIT 986
 122 GADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOOTRGLGCIIT 181
 987 GADTAACGIIISGLPVSARREIILGPADNPEGQWRLAPITAYSOOTRGLGCIIT 1046
 182 LTRGRDKQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 241
 1047 LTRGRDKQVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKPIITOMTND 1106
 242 DLVGMQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 301
 1107 DLVGMQAPPGASMPCTCGSSDLYLVTRHADVIIVRRGDSRGLSPRPVSYLKSSG 1166
 302 GGLLCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTR 341
 1167 GGLLCPSGHAGVIFRAAVCTRGVAAVDFIPVESMETTR 1206

RESULT 6

GNMVC

genome polyprotein - hepatitis C virus (strain HCV-1)
 N;Contains: capsid protein C, envelope protein M, hepatitis C virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
 C;Species: hepatitis C virus
 C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
 A;Accession: A39166; P04003; P04004
 R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coit Proc. Natl. Acad. Sci. U.S.A. 89, 2451-2455, 1991
 A;Title: Genetic organization and diversity of the hepatitis C virus.
 A;Reference number: A39166; MUID:91172826; PMID:1848704
 A;Accession: A39166

A:Molecule type: mRNA
 A:Residues: 13011 <CHO>
 A:Cross-references: GB:062221; NID:9329873; PIDN:AAA5676.1; PID:9329874
 R:Chan, S.W., McOmish, F., Holmes, E.C., Dow, B., Peuteher, U.F., Follett, E., Yap, P.L.
 J. Gen. Virol. 72, 1131-1141, 1992
 A>Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A:Reference number: PQ0393; MUID:92268871; PMID:1316933
 A:Accession: PQ0403
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CHA>
 A:Cross-references: DDBJ:D10128
 A:Experimental source: isolates E-b16
 A:Accession: PQ0404
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: isolates E-b17
 C:Superfamily: hepatitis C virus genome polypeptide; glycoprotein; hydrolase; nonstructura
 C:Keywords: ATP; capsid protein; envelope protein; predicted <CPC>
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein M #status predicted <EMP>
 F:192-389/Product: major envelope protein E #status predicted <NEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,446,476,532,540,556,576,623,645,1213,1255,2041,2077,22
 Query Match 88.0%; Score 1565; DB 1; Length 3011;
 Best Local Similarity 84.2%; Pred. No. 1,1e-123;
 Matches 287; Conservative 28; Mismatches 26; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLLAFGLMLVAGITVPPVPAQGLIRACMLV 60
 DB 866 VRGRDAIILITCAVHPELFDITKLLAFGLMLVAGITVPPVPAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAEFKALNLTGYVYDHLTPLODMAHAGRLDLAAVEPVIFSDMEVKIT 120
 DB 926 RKAAGHYVQMAEFKALNLTGYVYDHLTPLODMAHAGRLDLAAVEPVIFSDMEVKIT 985
 QY 121 WGAADTAACGDIISGLPVASARGREILIGPADNFEQCGKRLAPITAYSCQTRGLGCIIT 180
 DB 986 WGAADTAACGDIISGLPVASARGREILIGPADNFEQCGKRLAPITAYSCQTRGLGCIIT 1045
 QY 161 SLTGRDNQVGEVQVISTATQSFATCNGVCMVTFHAGSKTLGPKPIQMTYND 240
 DB 1046 SLTGRDNQVGEVQVISTATQSFATCNGVCMVTFHAGSKTLGPKPIQMTYND 1105
 QY 241 QDLVGMQAPPGASMTPTCGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 300
 DB 1106 QDLVGMQAPPGASMTPTCGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 1165
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVSMETMR 341
 DB 1166 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVSMETMR 1206

RESULT 7
 S40770 genome polypeptide - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: S40770; P01285
 R:Okamoto, H.
 Submitted to the EMBL Data Library, March 1992
 A:Reference number: S40770

A:Accession: S40770
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <OKA>
 A:Cross-references: EMBL:010749; NID:9221586; PIDN:BA01582.1; PID:9221587
 R:Okamoto, H., Okada, S., Sugiyama, Y., Yotsumoto, S., Tanaka, T., Yoshizawa, H., Tsuda,
 Jpn. J. Exp. Med. 60, 167-177, 1990
 A>Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A:Reference number: P01284; MUID:91013116; PMID:2170712
 A:Accession: P01285
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OK2>
 A:Cross-references: GB:P00831; NID:9221511; PIDN:BA00705.1; PID:9221512
 A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: major envelope protein M #status predicted <EMP>
 F:192-389/Product: major envelope protein E #status predicted <NEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 Query Match 87.7%; Score 1560; DB 1; Length 3011;
 Best Local Similarity 83.9%; Pred. No. 2.9e-123;
 Matches 286; Conservative 27; Mismatches 28; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLLAFGLMLVAGITVPPVPAQGLIRACMLV 60
 DB 866 VRGRDAIILITCAVHPELFDITKLLAFGLMLVAGITVPPVPAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAEFKALNLTGYVYDHLTPLODMAHAGRLDLAAVEPVIFSDMEVKIT 120
 DB 926 RKAAGHYVQMAEFKALNLTGYVYDHLTPLODMAHAGRLDLAAVEPVIFSDMEVKIT 985
 QY 121 WGAADTAACGDIISGLPVASARGREILIGPADNFEQCGKRLAPITAYSCQTRGLGCIIT 180
 DB 986 WGAADTAACGDIISGLPVASARGREILIGPADNFEQCGKRLAPITAYSCQTRGLGCIIT 1045
 QY 161 SLTGRDNQVGEVQVISTATQSFATCNGVCMVTFHAGSKTLGPKPIQMTYND 240
 DB 1046 SLTGRDNQVGEVQVISTATQSFATCNGVCMVTFHAGSKTLGPKPIQMTYND 1105
 QY 241 QDLVGMQAPPGASMTPTCGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 300
 DB 1106 QDLVGMQAPPGASMTPTCGSSDLVLYTRHADVIPIVRRDSDSGSLSPRPVSYLKSS 1165
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVSMETMR 341
 DB 1166 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVSMETMR 1206

RESULT 8
 G40770 genome polypeptide - hepatitis C virus (strain H)
 N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
 C:Species: hepatitis C virus
 A>Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Chan, S.W., McOmish, F., Holmes, E.C., Dow, B., Peuteher, U.F., Follett, E., Yap, P.L.
 Submitted to GenBank, July 1992
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>

A:Cross-references: GB:M67463; NID:g329737; PID:AAA5534.1; PID:g329738
 R:Inchasp, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Naeoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparat
 A:Reference number: A41546; MUID:92052256; PMID:1658800
 A:Contents: annotation
 A>Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
 F:1616-1866/Product: nonstructural protein NS4 #status predicted <NS4>
 F:1867-2017/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:156-209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1235,2041,2240,23

Query Match 87.5%; Score 1555; DB 1; Length 3011;
 Best Local Similarity 83.9%; Pred. No. 7,66-123;
 Matches 286; Conservative 29; Mismatches 26; Indels 0; Gaps 0;

QY 1 VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 60
 Db VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 60
 QY 866 VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 925
 Db VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 925
 QY 61 RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 120
 Db RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 120
 QY 926 RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 985
 Db RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 985
 QY 121 WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 180
 Db WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 180
 QY 986 WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 1045
 Db WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 1045
 QY 181 SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 240
 Db SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 240
 QY 1046 SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 1105
 Db SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 1105
 QY 241 QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 300
 Db QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 300
 QY 1106 QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 1165
 Db QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 1165
 QY 301 GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 341
 Db GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 341
 QY 1166 GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 1206
 Db GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 1206

RESULT 9
 JCS620
 genome polyprotein - hepatitis C virus (isolate EUH1460)
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: JCS620
 R:Chamborlin, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997
 A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant
 A:Reference number: JCS620; MUID:97365593; PMID:9223423
 A:Accession: JCS620
 A:Molecule type: mRNA
 A:Residues: 1-3014 <CHA>
 A:Cross-references: GB:Y13184
 A:Experimental source: genotype 5a, which predominates in South Africa
 A>Note: the translation of the nucleotide sequence is not complete in this paper
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:384-408/Product: hypervariable #status predicted
 F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1008-1616/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1231-1238/Product: nucleotide-binding motif A (P-loop)
 F:1313-1318/Product: nucleotide-binding motif B
 F:1317-1320/Product: DEXH motif
 F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4>
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2210-2249/Product: interferon sensitivity determining #status predicted

Query Match 76.6%; Score 1362; DB 1; Length 3014;
 Best Local Similarity 71.3%; Pred. No. 1,76-106;
 Matches 243; Conservative 49; Mismatches 49; Indels 0; Gaps 0;

QY 1 VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 60
 Db VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 60
 QY 867 VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 926
 Db VRGGRDAILITTCVHPELITFDITKLLAFGLPMLVQAGITKPYVRAQGLIRACMLV 926
 QY 61 RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 120
 Db RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 120
 QY 927 RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 986
 Db RKAAGHYVQVAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAEPVIFSDMEVKIT 986
 QY 121 WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 180
 Db WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 180
 QY 987 WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 1046
 Db WADTAACGDIISGLPVARSRRREILGPADNFEQGMRLIAPRYAQQOTGILGCIIT 1046
 QY 181 SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 240
 Db SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 240
 QY 1047 SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 1106
 Db SLTGRDNQVEGEVQVSTATQSFATCVNGCMTVFHAGSKTLAEPKPIQMTYND 1106
 QY 241 QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 300
 Db QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 300
 QY 1107 QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 1166
 Db QDLVWQAPPGARSMTPCTCGSSDLYVTRRADVYVARRGDSKSLSPRVSYLKSS 1166
 QY 301 GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 341
 Db GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 341
 QY 1167 GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 1207
 Db GGPIICPSGHAIVGFRAAVCTRGVAKAVDFIVESMETMR 1207

RESULT 10
 JQ1303
 genome polyprotein - hepatitis C virus (isolate HC-96)
 N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
 C:Accession: JQ1303
 R:Okamoto, H.; Okada, S.; Sugitani, Y.; Kural, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;
 J. Gen. Virol. 72, 2697-2704, 1991
 A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a huma
 A:Reference number: JQ1303; MUID:92044440; PMID:1658196
 A:Accession: JQ1303
 A:Molecule type: genomic RNA
 A:Residues: 1-3033 <OKA>
 A:Cross-references: GB:D00944; NID:g221650; PID:BA00792.1; PID:g221651
 A:Experimental source: isolate HC-96 from a Japanese individual
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transme
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEB>
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1011-1619/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1316-1321/Product: nucleotide-binding motif A (P-loop)
 F:1320-1323/Product: DEXH motif
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28
 Query Match 71.4%; Score 1269; DB 1; Length 3033;
 Best Local Similarity 67.2%; Pred. No. 1.3e-98;
 Matches 225; Conservative 49; Mismatches 63; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLLAIFGLMVLQAGITKVPYFVRAQGLIRACMLV 60
 DB VRGRDIIIIWAVALFCGVVFDITKMLLAVLGPAYLLKGLITVPYFVRHALLRWCTAV 929
 QY 61 RKAGGAYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAEPVFSDEMEKIT 120
 DB 930 RHLAGGAYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAEPVFSDEMEKIT 989
 QY 121 WGAADTAACGDIISGLPVASARGREILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
 DB 990 WGAETVACGDIILHGLPVASARGREVLGPADGYTSKMSKLAPITAYSOQTRGLGCIIV 1049
 QY 181 SLTGRDNQVGEVQVNSTATQSLFATCVNGVCMVTHGAGSKTLGPKPFIQMTYND 240
 DB 1050 SLTGRDNQVGEVQVNSTATQSLFATCVNGVCMVTHGAGSKTLGPKPFIQMTYNSAE 1109
 QY 241 QDLVQMAFPGASMTCTCGSSDLYLVTNRADYIPVRKRDGSGSLSPRPVSYLKSS 300
 DB 1110 QDLVQMAFPGASMTCTCGSSDLYLVTNRADYIPVRKRDGSGSLSPRPVSYLKSS 1169
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPIPVESMETTR 341
 DB 1170 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPIPVESMETTR 1210

RESULT 11

genome polypeptide - hepatitis C virus (strain HC-78)
 N:Contains: capsid protein C; envelope protein M; hepatitis virus (HCV 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40250; PMID:1314459
 R:Okamoto, H.; Kunita, K.; Okada, S.I.; Yamamoto, K.; Mizuka, H.; Tanaka, T.; Fukuda, S.;
 Virology 188, 331-341, 1992
 A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep
 A:Reference number: A40250; MUID:92230232; PMID:1314459
 A:Accession: A40250
 A:Molecule type: genomic RNA
 A:Residues: 1-3033 <OKA>
 A:Cross-references: GB:D10988; GB:D01221; NID:9221608; PIDN:BA01761.1; PID:9221609
 R:Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
 J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A:Reference number: P00393; MUID:92268871; PMID:1316939
 A:Accession: P00393
 A:Molecule type: genomic RNA
 A:Residues: 2678-2754 <CHA>
 A:Cross-references: DDBJ:D10134
 A:Experimental source: isolate E-b12
 R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991
 A:Title: Distribution of plural HCV types in Japan.
 A:Reference number: P00554; MUID:92068204; PMID:11720309
 A:Accession: P00554
 A:Molecule type: mRNA
 A:Residues: 2678-2729 <KAT>
 A:Cross-references: GB:D10562; GB:D90518; NID:9221523; PIDN:BA01418.1; PID:9221524
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 P:1-115/Product: capsid protein C #status predicted <CPC>
 P:116-181/Product: envelope protein M #status predicted <EPM>
 P:192-389/Product: major envelope protein E #status predicted <MEP>
 P:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
 P:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 P:1011-1619/Product: hepatitis virus #status predicted <NS3>
 P:11234-1241/Region: nucleotide-binding motif A (P-loop)

F:1316-1321/Region: nucleotide-binding motif B
 F:1320-1323/Region: DEKH motif
 F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4a>
 F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>
 F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,231
 Query Match 71.0%; Score 1262; DB 1; Length 3033;
 Best Local Similarity 66.0%; Pred. No. 4.9e-99;
 Matches 225; Conservative 54; Mismatches 62; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHPELFDITKLLAIFGLMVLQAGITKVPYFVRAQGLIRACMLV 60
 DB VRGRDIIIIWAVALFCGVVFDITKMLLAVLGPAYLLKGLITVPYFVRHALLRWCTAV 929
 QY 61 RKAGGAYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAEPVFSDEMEKIT 120
 DB 930 RHLAGGAYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAEPVFSDEMEKIT 989
 QY 121 WGAADTAACGDIISGLPVASARGREILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
 DB 990 WGAETVACGDIILHGLPVASARGREVLGPADGYTSKMSKLAPITAYSOQTRGLGCIIV 1049
 QY 181 SLTGRDNQVGEVQVNSTATQSLFATCVNGVCMVTHGAGSKTLGPKPFIQMTYND 240
 DB 1050 SLTGRDNQVGEVQVNSTATQSLFATCVNGVCMVTHGAGSKTLGPKPFIQMTYNSAE 1109
 QY 241 QDLVQMAFPGASMTCTCGSSDLYLVTNRADYIPVRKRDGSGSLSPRPVSYLKSS 300
 DB 1110 QDLVQMAFPGASMTCTCGSSDLYLVTNRADYIPVRKRDGSGSLSPRPVSYLKSS 1169
 QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPIPVESMETTR 341
 DB 1170 GGPLCPSGHAAGIFRAAVCTRGVAKAVDPIPVESMETTR 1210

RESULT 12

polypeptide - douroucouli hepatitis GB virus A
 C:Species: douroucouli hepatitis GB virus A
 C:Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text_change 17-Nov-2000
 C:Accession: T08841
 R:Berker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Morice, C.C.; Mushahar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A:Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A:Reference number: Z16486; MUID:98120818; PMID:9460920
 A:Accession: T08841
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-3005 <EMK>
 A:Cross-references: EMBL:A023425; NID:92828599; PIDN:AA040502.1; PID:92828600
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: polypeptide

Query Match 23.0%; Score 408.5; DB 2; Length 3005;
 Best Local Similarity 31.3%; Pred. No. 8.9e-26;
 Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;
 QY 2 RGRD-----AIIITCAVHPELFDITKLLAIFGLMVLQAGITKVPYFVRAQGLIRAC 57
 DB 813 RGRDWRVTAWVAAGIIFPREVVCSPAALTA-----LAALDISIDYLETL-ILTPA 864
 QY 58 MLVRKA-----AGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAEPV 109
 DB 865 QPARARLLDSITFLGDDLTFRFAVRRLERGVTLFOHCGOVSXGAALILDLGVALEPV 924
 QY 110 IFSDMEVKIITWGAADTAACGDIISGLPVASARGREILG--PADNFEQGWRLAPITAY 167
 DB 925 SVTAADCVIYHAAATTLAQGVVELPVVARGBDVLGVPPSVRALPPGVPYPAVVV- 983
 QY 166 SQQTGLGICITTSLTGRDNQVGEVQVNSTATQSLFATCVNGVCMVTHGAGSKTLG 227
 DB 984 MORGGLFSPSVKTSMLGRDREHESIVLIGITTSRSVGTGVNGVMTTFHGSARKTLAG 1043


```

QY 15 VHPRL-----FDITKLLATFSPMLVQAGI-----TKVPFVRAQGLIRACMYR 61
Db 154 VHDVLVLMGGEVSELOPAKSPVILLAGLVGKTTVCATLACTYKKG--KSCMLI- 210
QY 62 KAAAGHYVQMAFMKLAAL---TGIVYVDHLTPLO--DWAHAGLRDLAVAVEPVI FSDMEV 116
Db 211 --AGDVYRPAIDQLVLGEQVGVVYTAGTDVKPADIAKQGLKEAK-----NNVDV 261
QY 117 KIITWGADTACGDIISGLFVSARGREL-----LGPADNFEQGWRLAPITAYSQGT 171
Db 262 VIM---DTAGRLQIDKGMDELKDVKKFLNPTFVLLVDMTGO--EAAALVTTFNVEI 315
QY 172 RGLIGCIITSUTGRDNQVEGEVQVVS 198
Db 316 -GITGALITKLDGDSRGAALSVKEYS 341

```

Search completed: May 6, 2004, 09:37:20
 Job time : 12.1056 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 7.119 Seconds
(without alignments)

2494.160 Million cell updates/sec

Title: US-10-650-585-14

Perfect score: 1778
1 VRGGRDAIILLTCVAPHELI.....RGVAKAVDIPVSEMETTMR 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1711	96.2	3010	POLG_HCVUT
2	1710	96.2	3010	POLG_HCVUT
3	1685	94.8	3010	POLG_HCVUT
4	1649	92.7	3010	POLG_HCVUT
5	1565	88.0	3011	POLG_HCV1
6	1555	87.5	3011	POLG_HCV1
7	1269	71.4	3033	POLG_HCV6
8	1262	71.0	3033	POLG_HCV8
9	102.5	5.8	660	VST2_HEVBU
10	102.5	5.8	660	VST2_HEVBU
11	101	5.7	564	SR5C_ARATH
12	101	5.7	600	PRO2_MOUSE
13	93.5	5.3	1380	CYRA_LEIDO
14	93	5.2	434	TOLB_CHLIE
15	92.5	5.2	706	TRFE_HORSE
16	92	5.2	659	VST2_HEVME
17	91.5	5.1	660	VST2_HEVME
18	91.5	5.1	660	VST2_HEVME
19	91.5	5.1	3414	POLG_TBEVH
20	90.5	5.1	3414	POLG_TBEVH
21	89.5	5.0	961	ATCU_YERPE
22	87.5	4.9	3412	POLG_TBEVH
23	86.5	4.9	347	MDHM_EUCGU
24	86.5	4.9	470	MDHM_EUCGU
25	85	4.8	338	GATE_NERGO
26	85	4.8	433	DCUA_MOLSU
27	85	4.8	470	NRAM_IATRA
28	85	4.8	730	HELS_METVA
29	85	4.8	854	PWP2_SCHPO
30	85	4.8	3313	CLR3_RAT
31	84.5	4.8	1705	PTPV_MOUSE
32	84	4.7	309	UCP2_RAT
33	84	4.7	339	GPDA_CORBF

34	84	4.7	423	1	MR09_MOUSE	Q9wtu6 mus musculus
35	84	4.7	470	1	NRAM_IARUE	P31510 influenza a
36	84	4.7	503	1	YAO2_MYCTU	O05586 mycobacteri
37	83.5	4.7	538	1	DAC_ACTSP	P39045 actinomadar
38	83.5	4.7	1399	1	PROG_PSEAR	Q9hw9 pseudomonas
39	83	4.7	341	1	MDHM_BRANA	O43744 brassica na
40	82.5	4.6	453	1	NRAM_TAVIL	P03470 influenza a
41	82.5	4.6	711	1	HPA2_RHIME	Q8vzr3 rhizodium m
42	82	4.6	309	1	UCP2_MOUSE	P70406 mus musculus
43	82	4.6	384	1	MR08_BRARE	Q9gd9s brachydanio
44	82	4.6	403	1	PGK_CHIMU	O9pln4 chlamydia m
45	82	4.6	425	1	VIBR_RAT	P48974 rattus norv

ALIGNMENTS

RESULT 1	ID	SEQUENCE	STANDARD	PRT: 3010 AA.
POLG_HCVUT	AC	Q00269;		
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].			
DE	Hepatitis C virus (isolate HC-JT) (HCV).			
OS	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitisin.			
OC	NCBI_TaxID=31642;			
OX	NCBI_TaxID=31642;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92295714; PubMed=1318627;			
RA	Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.,			
RT	Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals."			
RT	Infected individuals."			
RT	Virus Res. 23:39-53(1992).			
CC	-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.			
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).			
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and RNA.			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL, D11168; AAA01943.1; -			
DR	PIR, A45573; A45573.			
DR	MEROPS, S29.001; -			
DR	MEROPS, U39.001; -			
DR	InterPro, IPR009003; Cys_Ser_trypsin.			
DR	InterPro, IPR001410; DEAD.			
DR	InterPro, IPR002522; HCV_capsid.			

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D90208; BAA14233.1; -
 DR PIR; A39253; GNVVCT.
 DR HSSP; P26663; LXP.
 DR MEROPS; S29.001; -
 DR MEROPS; U39.001; -
 DR InterPro; IPR009033; Cys_Ser_trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVLr.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR Pfam; PF01606; HCV_NS1; 1.
 DR Pfam; PF01606; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KM Transmembrane; Nonstructural
 INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT STATE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2768 2768 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3010 AA; 327017 KM; AA993794P6DB185 CR664;
 SQ SEQUENCE
 Query Match 96.2%; Score 1710; DB 1; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 1.7e-136;
 Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;
 QY 1 VRGGRDAITLLTCVAPPELIPDITLTLAIGPPLVLTAGITKVPYFRAQGLRACMLV 60
 DB 866 VRGGRDAITLLTCVAPPELIPDITLTLAIGPPLVLTAGITKVPYFRAQGLRACMLV 925
 QY 61 RRAAGHYVQVAFMKLAALGTGYVYDHTPLQDWAHAGLRDLAVAVEPVSDEWEKITT 120
 DB 926 RRAAGHYVQVAFMKLAALGTGYVYDHTPLQDWAHAGLRDLAVAVEPVSDEWEKITT 985
 QY 121 WADPTAACGDIISGLPVARARGRETLGPADNPEQGRLLAPITAYSOQTRGLGCIIT 180
 DB 986 WADPTAACGDIISGLPVARARGRETLGPADNPEQGRLLAPITAYSOQTRGLGCIIT 1045
 QY 181 SLTGDRKNQVEGEVQVNSTATQSFATCNGVCWTFPHGASXTLAGPKGPIYQYTNVD 240
 DB 1046 SLTGDRKNQVEGEVQVNSTATQSFATCNGVCWTFPHGASXTLAGPKGPIYQYTNVD 1105
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYTRHADVTPVRRGDSRGLSPRVSTLKSS 300
 DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLVYTRHADVTPVRRGDSRGLSPRVSTLKSS 1165
 QY 301 GGELICPSGHAIGFRAVCTRGAKAVDFIPVESMETTMR 341
 DB 1166 GGELICPSGHAIGFRAVCTRGAKAVDFIPVESMETTMR 1206
 RESULT 3
 POLG_HCVTM STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (rel. 25, Created)
 DT 01-APR-1993 (rel. 25, Last sequence update)
 DT 10-OCT-2003 (rel. 42, Last annotation update)
 DE Genome polypeptide [contains: Capsid protein C (Core protein) (P222);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)1].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus;
 OX NCBI_TaxID=31645;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' terminal of viral genome and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

	Query Match	Best Local Similarity	Score	DB 1	Length	DB 2	Score	DB 3	Length	DB 4	Score	DB 5	Length	DB 6	Score	DB 7	Length	DB 8	Score	DB 9	Length	DB 10	Score	DB 11	Length	DB 12	Score	DB 13	Length	DB 14	Score	DB 15	Length	DB 16	Score	DB 17	Length	DB 18	Score	DB 19	Length	DB 20	Score	DB 21	Length	DB 22	Score	DB 23	Length	DB 24	Score	DB 25	Length	DB 26	Score	DB 27	Length	DB 28	Score	DB 29	Length	DB 30	Score	DB 31	Length	DB 32	Score	DB 33	Length	DB 34	Score	DB 35	Length	DB 36	Score	DB 37	Length	DB 38	Score	DB 39	Length	DB 40	Score	DB 41	Length	DB 42	Score	DB 43	Length	DB 44	Score	DB 45	Length	DB 46	Score	DB 47	Length	DB 48	Score	DB 49	Length	DB 50	Score	DB 51	Length	DB 52	Score	DB 53	Length	DB 54	Score	DB 55	Length	DB 56	Score	DB 57	Length	DB 58	Score	DB 59	Length	DB 60	Score	DB 61	Length	DB 62	Score	DB 63	Length	DB 64	Score	DB 65	Length	DB 66	Score	DB 67	Length	DB 68	Score	DB 69	Length	DB 70	Score	DB 71	Length	DB 72	Score	DB 73	Length	DB 74	Score	DB 75	Length	DB 76	Score	DB 77	Length	DB 78	Score	DB 79	Length	DB 80	Score	DB 81	Length	DB 82	Score	DB 83	Length	DB 84	Score	DB 85	Length	DB 86	Score	DB 87	Length	DB 88	Score	DB 89	Length	DB 90	Score	DB 91	Length	DB 92	Score	DB 93	Length	DB 94	Score	DB 95	Length	DB 96	Score	DB 97	Length	DB 98	Score	DB 99	Length	DB 100	Score	DB 101	Length	DB 102	Score	DB 103	Length	DB 104	Score	DB 105	Length	DB 106	Score	DB 107	Length	DB 108	Score	DB 109	Length	DB 110	Score	DB 111	Length	DB 112	Score	DB 113	Length	DB 114	Score	DB 115	Length	DB 116	Score	DB 117	Length	DB 118	Score	DB 119	Length	DB 120	Score	DB 121	Length	DB 122	Score	DB 123	Length	DB 124	Score	DB 125	Length	DB 126	Score	DB 127	Length	DB 128	Score	DB 129	Length	DB 130	Score	DB 131	Length	DB 132	Score	DB 133	Length	DB 134	Score	DB 135	Length	DB 136	Score	DB 137	Length	DB 138	Score	DB 139	Length	DB 140	Score	DB 141	Length	DB 142	Score	DB 143	Length	DB 144	Score	DB 145	Length	DB 146	Score	DB 147	Length	DB 148	Score	DB 149	Length	DB 150	Score	DB 151	Length	DB 152	Score	DB 153	Length	DB 154	Score	DB 155	Length	DB 156	Score	DB 157	Length	DB 158	Score	DB 159	Length	DB 160	Score	DB 161	Length	DB 162	Score	DB 163	Length	DB 164	Score	DB 165	Length	DB 166	Score	DB 167	Length	DB 168	Score	DB 169	Length	DB 170	Score	DB 171	Length	DB 172	Score	DB 173	Length	DB 174	Score	DB 175	Length	DB 176	Score	DB 177	Length	DB 178	Score	DB 179	Length	DB 180	Score	DB 181	Length	DB 182	Score	DB 183	Length	DB 184	Score	DB 185	Length	DB 186	Score	DB 187	Length	DB 188	Score	DB 189	Length	DB 190	Score	DB 191	Length	DB 192	Score	DB 193	Length	DB 194	Score	DB 195	Length	DB 196	Score	DB 197	Length	DB 198	Score	DB 199	Length	DB 200	Score	DB 201	Length	DB 202	Score	DB 203	Length	DB 204	Score	DB 205	Length	DB 206	Score	DB 207	Length	DB 208	Score	DB 209	Length	DB 210	Score	DB 211	Length	DB 212	Score	DB 213	Length	DB 214	Score	DB 215	Length	DB 216	Score	DB 217	Length	DB 218	Score	DB 219	Length	DB 220	Score	DB 221	Length	DB 222	Score	DB 223	Length	DB 224	Score	DB 225	Length	DB 226	Score	DB 227	Length	DB 228	Score	DB 229	Length	DB 230	Score	DB 231	Length	DB 232	Score	DB 233	Length	DB 234	Score	DB 235	Length	DB 236	Score	DB 237	Length	DB 238	Score	DB 239	Length	DB 240	Score	DB 241	Length	DB 242	Score	DB 243	Length	DB 244	Score	DB 245	Length	DB 246	Score	DB 247	Length	DB 248	Score	DB 249	Length	DB 250	Score	DB 251	Length	DB 252	Score	DB 253	Length	DB 254	Score	DB 255	Length	DB 256	Score	DB 257	Length	DB 258	Score	DB 259	Length	DB 260	Score	DB 26
--	-------------	-----------------------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	-------

FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65f8c9447fcb5af9 CRC64;

Query Match 88.0%; Score 1565; DB 1; Length 3011;
 Best Local Similarity 84.2%; Pred. No. 3.3e-124;
 Matches 287; Conservative 28; Mismatches 26; Indels 0; Gaps 0;

QY 1 VRGGRDAIITLTCAVHELIPTDKLLATFGPLMTQAGITVPPVPRVQGLIRACMV 60
 DB 866 VRGRDVIILMCAVHPTLVFDITKLLAVFGPLMTQASLKVPRVQGLIRFCALA 925
 QY 61 RKAGGHVQMAEFKALATGVVYDHLTPLODMAHAGLRDLAVAPVTFSDMEVKIT 120
 DB 926 RKMIQGHVQVITKLGALGTYYNNLTPLRDANHGLMDLAVAPVTFSDMEVKIT 985
 QY 121 WGAIDTAACGDIISLPLVSARGREILLGPADNFGCGMRLAPITANSQOTRGLGCIIT 180
 DB 986 WGAIDTAACGDIISLPLVSARGREILLGPADNFGCGMRLAPITANSQOTRGLGCIIT 1045
 QY 181 SLTGRDNQVEGEVQVNSTATOSFLATCVGCVCTVHGAGSKTLAPKXPITOMTND 240
 DB 1046 SLTGRDNQVEGEVQVNSTATOSFLATCVGCVCTVHGAGSKTLAPKXPITOMTND 1105
 QY 241 QDLVWQAPPGARSMPTCTCGSSDLYIVTRHADVIVRRRGDSRGLSPVSYLKSS 300
 DB 1106 QDLVWQAPPGARSMPTCTCGSSDLYIVTRHADVIVRRRGDSRGLSPVSYLKSS 1165
 QY 301 GGPIICSGHVGIFRAVCTRGVAKAVDFIPVSMETMR 341
 DB 1166 GGPIICSGHVGIFRAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 6
 POLG HCVM STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepatitisin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus
 CX NCBI_TaxID=11108;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebadee S., Lee D.H.H., Sugtani M., Nasoff M.,
 RA Prince A.M.,
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.",
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.",
 RL Nat. Struct. Biol. 4:463-467 (1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.,
 RT Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.",
 RL Structure 6:89-100 (1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the pe
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL, M67463; AAAS534.1; -.
 DR PIR, A36814; GNWVCH.
 DR PDB, 1HEI; 25-NOV-98.
 DR PDB, 1AIV; 16-FEB-99.
 DR PDB, 1AIR; 17-JUN-98.
 DR MEROPS, S29.001; -.
 DR TRANSFAC, T04155; -.
 DR InterPro, IPR009003; Cys_Ser_trypsin.
 DR InterPro, IPR001410; DEAD.
 DR InterPro, IPR002522; HCV capsid.
 DR InterPro, IPR002521; HCV core.
 DR InterPro, IPR002519; HCV env.
 DR InterPro, IPR002531; HCV NS1.
 DR InterPro, IPR002518; HCV NS2.
 DR InterPro, IPR000745; HCV NS4A.
 DR InterPro, IPR001490; HCV NS4B.
 DR InterPro, IPR002868; HCV NS5A.
 DR InterPro, IPR002166; HCV RdRp.
 DR InterPro, IPR001650; Helicase_C.
 DR InterPro, IPR004109; Peptidase_C29.
 DR InterPro, IPR007095; RNA_pol_DS_PS.
 DR InterPro, IPR007094; RNA_pol_PSVir.
 DR Pfam, PF01543; HCV capsid; 1.
 DR Pfam, PF01539; HCV env; 1.
 DR Pfam, PF01542; HCV core; 1.
 DR Pfam, PF01560; HCV NS1; 1.
 DR Pfam, PF01538; HCV NS2; 1.
 DR Pfam, PF02807; HCV NS3; 1.
 DR Pfam, PF01006; HCV NS4a; 1.
 DR Pfam, PF01001; HCV NS4b; 1.

FT	TURN	1419	1420
FT	STRAND	1432	1436
FT	TURN	1438	1439
FT	STRAND	1450	1453
FT	STRAND	1456	1463
FT	STRAND	1471	1478
FT	STRAND	1480	1480
FT	HELI-X	1481	1488
FT	TURN	1489	1490
FT	STRAND	1497	1501
FT	STRAND	1507	1507
FT	STRAND	1511	1511
FT	HELI-X	1514	1527
FT	HELI-X	1532	1544
FT	STRAND	1550	1550
FT	HELI-X	1555	1564
FT	HELI-X	1570	1578
FT	TURN	1579	1580
FT	HELI-X	1584	1597
FT	TURN	1598	1598
FT	HELI-X	1606	1611
FT	TURN	1614	1618
FT	STRAND	1622	1623
FT	STRAND	1627	1627
FT	STRAND	1635	1636
FT	HELI-X	1640	1652
SQ	SEQUENCE	3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;	
Query Match	Best Local Similarity	87.5%; Score 1555; DB 1; Length 3011;	
Matches	286; Conservative	29; Mismatches	26; Indels
Gaps	0;		

OY	1 VRGRDAIILITCAVHPELLFDITKLLIAIFGPLMTIOAGITKVPIFYRAOGLIRACMLV	60
Ds	866 VRGRDRIVLILTCTCVHPELVFEDITKLLIAIFGPLMIIGASILKKVPFVRVOGLRICALA	925
OY	61 PKACGHYVQMAEFKLALNGTYYDHLLTPLODMAHAGLDLAAYVEPVTFSDMEVKIIT	120
Ds	926 RKINGHHVQAIIKLGALTGTGVNHLAPURDMAHGLRDLAAYVEPVTFSRMETKIIT	985
OY	121 WGAADTAACGDIIISGLPYSAARRGREILLGPADNFEGOGRRLLAPIPAYVSQOTRGLLCIIT	180
Ds	986 WGAADTAACGDIIINGLPYSARRGEIILLGPADGWYSKGMRLLAIPITAYAQCOTRGLLCIIT	1045
OY	181 SLTRDKNOVEGEVQVYSTATOSLATCANGVCMTVFHGAGSKTLAHPKGPIOTMNTND	240
Ds	1046 SLTRDKNOVEGEVQIVSTATQTLEATCINGVCMTVFHGAGRITLASPKGPIOTYTND	1105
OY	241 QDLVGMOAPPARSMPTCTCGSSDLYIVYTRHADVIPIYRRAGDSKGSLLSPRPVSYLKSSS	300
Ds	1106 QDLVGWAPPOGSRSLTLETCOSSDLYIVYTRHADVIPYRRAGDSKGSLLSPRPISYLKSSS	1165
OY	301 GGPLICPSGHAIVGIFRAAVCTRGVAKAVDFIPVESNETTNR	341
Ds	1166 GGPLICTETGHAIVGFRAAVCTRGVAKAVDFIPVENLETITWR	1206

RESULT 7	
POLG_HCVU6	STANDARD; PRT; 3033 AA.
ID	POLG_HCVU6
AC	P26660;
DT	01-AUG-1992 (Rel. 23, Created)
DT	01-AUG-1992 (Rel. 23, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)]
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OC	Hepaditis C virus (isolate HC-J6) (HCV).
OC	Vitruises; ssRNA positive-strand vitruises, no DNA stage; Flaviviridae.

CC Hepatitis. NCI TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N. A.
 EX MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Iizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions";
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: The small proteins NS2a, NS2b, NS4a and NS4b are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and RNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL: D00944; BA00792.1; -
 DR PIR: J01303; J01303.
 DR HSP: P27958; 1HE1.
 DR MEROPS: S29.001; -
 DR
 DR InterPro: IPR009003; Cys_Ser_lysine.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C9.
 DR InterPro: IPR007095; RNA_pol_DS_P.
 DR InterPro: IPR007094; RNA_pol_PSVlt.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD16062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KM Transmembrane; Nonstructural
 FT INT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT ACT_SITE 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; P957F5C1A273BE9E CRC64;

Query Match 71.4%; Score 1269; DB 1; Length 3033;
 Best Local Similarity 67.2%; Pred. No. 4,2e-99;
 Matches 229; Conservative 49; Mismatches 63; Indels 0; Gaps 0;

QY 1 VREGPDAIILLCANHEPILFDITKLLAIFGELVLTQNGITKVPYFPAAGLIRACMLV 60
 DB 870 VREGPDAIILLCANHEPILFDITKLLAIFGELVLTQNGITKVPYFPAAGLIRACMLV 929
 QY 61 RKAAGHYQMAFMKALATGYVDHLTPLDQMAHAGRDAAVAPYIFSDMEKITT 120
 DB 930 RHLAAGRYQMLLALGRWTGYIYDHLTPMDMANGRDAVAPYIFSDMEKIV 989
 QY 121 WGAADTAACDIIISGLPVSARREBILLPADNPEQGGWMLAPITAYASQOTRGLGCIIT 180
 DB 990 WGAETACCDIIHGLPVSARLREVLILPADDYTSKMSLLAPITAYASQOTRGLGIIIV 1049
 QY 181 SLTGRPKNOVEGEVOVVSATQSPFLATCVNGCWVFEHAGSKTLAAGPITQWYTNVD 240
 DB 1050 SMTGRDKTEQAGEIVLSTVTSFGLTTISGLVMTVYHAGSKTLAAGPITQWYTNVD 1109
 QY 241 ODVLGWAQPPGARSMPTCTGSSDLYLVTRHADVLPVRRRGSRSGLSPFVSYLKSS 300
 DB 1110 GDLVGMPSPPGTXKSLPECTCGAVDLYLVTRNADVLPARRRGRKGLALSPRLSTLKSS 1169
 QY 301 GGPLCPGSHAVGIRPAACTGCAKADPTVESMETMR 341
 DB 1170 GGPLCPGSHAVGIRPAACTGCAKADPTVESMETMR 1210

RESULT 8
 POSG_HCVU8 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP6) (GP70) (NS1); Protein P7, Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis C virus).]

(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48).
 OS Hepatitis C virus (isolate Hc-J8) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; HCV.
 CC NCBI_TaxId=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9230233; Pubmed=1314459;
 RA Okamoto H., Kura K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes";
 RT Virology 188:333-341 (1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polypeptide, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.1sb-sib.ch/announce/> or send an email to license@1sb-sib.ch).
 CC -----
 CC EMBL; D10988; BAA01761.1; -;
 CC PIR; A40250; GNWVJ8.
 CC HSSP; P27959; 1HE1.
 CC MEROPS; S29.001; -;
 DR InterPro: IPR009003; Cys_Ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR00745; HCV_NS4A.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSTsr.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4A; 1.
 DR Pfam; PF01001; HCV_NS4B; 1.
 DR Pfam; PF01506; HCV_NS5A; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SMC0487; DEXDC; 1.
 DR PolProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KM Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1
 FT 1 REMOVED FROM CAPSID PROTEIN C BY THE

Query Match	Best Local Similarity	Score	DB 1	Length	3033
Query 225; Conservative	66.0%	Pred. No. 1.6e-98	Matches 62; Mismatches 62; Indels 0; Gaps 0;		
1 VGGADATILLTCAVPELIFDITKLLAIGPLVLOAGITKVPYFVAGLIRACMLV	60				
870 VRGGRDGIIMWAVILHPLVFEVTKMLAIIIGPAVILKASLRIPYFRAHALRVCTLV	929				
61 KKAAGHYVQAFMKLALITGTYYVDHITPIQDMAHAGLRLAVAVEVIESDMEVKIT	120				
930 KHLAAGARYIOMLITIGMTGYIYDHSPLSTAAAGLRDLAAVEVIESDMEVKIV	989				
121 WGAADTAAGDIIISGLPVASARRGREILIPADNFBGQGRILAPITAYSOQTRGLGCIIT	180				
990 WGAETVACGDILHGLPVASARRGREILIPADNFBGQGRILAPITAYSOQTRGLGCIIT	1049				
181 SLTGDKNQVEGEVAVSTAFQSLATVNGVCVTVFPGAGSKTLGKSGITOMTYND	240				
1050 SLTGRDKNEQAGVQVLSVQTFLGTSISGVLTAVVGAQNKTLGAGKGVITMYISAE	1109				
241 ODVGMQAPPGARSMTPCTGSSSDIYVTRADYIVPVRARSDSGSLSPVSYLKSS	300				
1110 GDLVGPSPGPKSLDPTCGAVDYIVTRADYIVPVRARSDSGSLSPVSYLKSS	1169				
301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSEMTTMR	341				
1170 GSPVLCPSRSHAVGLFRAAVCARGVAKSIDFIPVESLDVATR	1210				

RESULT 9
 VST2 HEVBU STANDARD; PRT; 660 AA.
 AC P29326; Rel. 24, Created
 DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Structural protein 2 precursor (ORF2)
 OS Hepatitis E virus (strain Burma) (HEV)
 CC Viruses; ssRNA positive-strand viruses, no DNA stage;
 CC Hepatitis E-like viruses.
 CC NCBI_TaxID=31767;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=92024067; PubMed=1926770;
 CC Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
 CC Fry K.E., Reyes G.R.;
 CC "Hepatitis E virus (HEV): molecular cloning and sequencing of the
 CC full-length viral genome";
 CC Virology 185:120-131(1991).
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M73218; AAA45736.1; -
 CC DR PIR: C40778; VHWMH2.
 CC DR InterPro: IPR004261; SP2.
 CC DR InterPro: IPR008975; Viral_cap_coat.
 CC DR Pfam: PF03014; SP2; 1.
 CC KW Signal.
 CC FT SIGNAL 1 19 POTENTIAL.
 CC FT CHAIN 20 660 STRUCTURAL PROTEIN 2.
 CC SQ SEQUENCE 660 AA; 70978 MM; 5832A013CCCA61C CRC64;
 CC
 CC Query Match 5.8%; Score 102.5; DB 1; Length 660;
 CC Best Local Similarity 19.8%; Pred. No. 0.55;
 CC Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;
 CC
 CC 24 TKLL--AIFGLMLVQAG-----ITKPYFVR---AAGLRACMLYKAGGHVQMA 72
 CC 151 TNLVLYAAPLSPLLPDQDNTNTHMATEASNAQRYRATIRYPLVNNAGVYALIS 210
 CC 73 FMKLAALGTYYVDHLTPLODWAHAGLDLAVAEVPIFSDMEVKIITWGADTAACGDI 132
 CC 211 FWPQTTTPTSV-----DNMSITSDVRLVPGIASELVI----- 246
 CC 133 SGLPVARSRRREILIGPAD--NFGQGWRLAP--TVSQQTRGL-----GCIITSLNG 184
 CC 247 -----PSERLHYRNQGMRSVETSGVAEEBATSGLVCLIHGSLVNSYTN 290
 CC 185 -----FDKNQVEGENVVVSTATQSF 205
 CC 291 TPTYGALGLDPALEFRNLTPGNTNTRVRSYSTARHRLRGADGTALITTAATRFM 350
 CC 206 A---TCVNGV-----CMTVFH-----GAG-----SKTLAPRG--PIT 233
 CC 351 KDLYFTSTNGVEIGRGIALTLFNLADTLGLPTLISAGGOLFVSRPVASANGEPTV 410
 CC 234 QMYTNVDQDLVGMQAPGARSMTPCTGSSDLVYV---TRHADVIPIVRRGDSRG--SLIS 289
 CC 411 KLTYSVENA---QODKGIALPHDIDLESRRVVIQDYDNQHEODRPTSPAPSRFFSVLR 466
 CC 290 PRPVSYLK-----GSSGGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPV 333
 CC 467 ANDVLMILSLTAAYDOSTYSGSTGPPYV--SDSVTLVNVATGAQAVASIDMTKV 519
 CC
 CC RESULT 10
 CC VST2_HEVPA
 CC ID VST2_HEVPA STANDARD; PRT; 660 AA.
 CC AC P33426;

DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Structural protein 2 precursor (ORF2)
 OS Hepatitis E virus (strain Pakistan) (HEV)
 CC Viruses; ssRNA positive-strand viruses, no DNA stage;
 CC Hepatitis E-like viruses.
 CC NCBI_TaxID=33774;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=92115700; PubMed=1713327;
 CC Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
 CC Malik I.A., Iqbal M., Purcell R.H.;
 CC "Characterization of a prototype strain of hepatitis E virus";
 CC Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
 CC -1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
 CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
 CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M80581; AAA5727.1; -
 CC DR InterPro: IPR004261; SP2.
 CC DR InterPro: IPR008975; Viral_cap_coat.
 CC DR Pfam: PF03014; SP2; 1.
 CC KW Signal.
 CC FT SIGNAL 1 22 BY SIMILARITY.
 CC FT CHAIN 23 660 STRUCTURAL PROTEIN 2.
 CC SQ SEQUENCE 660 AA; 70980 MM; 8085BC3C7B46FD3 CRC64;
 CC
 CC Query Match 5.8%; Score 102.5; DB 1; Length 660;
 CC Best Local Similarity 19.8%; Pred. No. 0.55;
 CC Matches 82; Conservative 54; Mismatches 128; Indels 151; Gaps 20;
 CC
 CC 24 TKLL--AIFGLMLVQAG-----ITKPYFVR---AAGLRACMLYKAGGHVQMA 72
 CC 151 TNLVLYAAPLSPLLPDQDNTNTHMATEASNAQRYRATIRYPLVNNAGVYALIS 210
 CC 73 FMKLAALGTYYVDHLTPLODWAHAGLDLAVAEVPIFSDMEVKIITWGADTAACGDI 132
 CC 211 FWPQTTTPTSV-----DNMSITSDVRLVPGIASELVI----- 246
 CC 133 SGLPVARSRRREILIGPAD--NFGQGWRLAP--TVSQQTRGL-----APITAYSQ 169
 CC 247 -----PSERLHYRNQGMRSVETSGVAEEBATSGLVCLIHGSLVNSYTN 290
 CC 170 QT-RGLGCI-----ITSLTGRDKNQ-----VEGEVQVSTATQSF 205
 CC 291 TPTYGALGLDPALEFRNLTPGNTNTRVRSYSTARHRLRGADGTALITTAATRFM 350
 CC 206 A---TCVNGV-----CMTVFH-----GAG-----SKTLAPRG--PIT 233
 CC 351 KDLYFTSTNGVEIGRGIALTLFNLADTLGLPTLISAGGOLFVSRPVASANGEPTV 410
 CC 234 QMYTNVDQDLVGMQAPGARSMTPCTGSSDLVYV---TRHADVIPIVRRGDSRG--SLIS 289
 CC 411 KLTYSVENA---QODKGIALPHDIDLESRRVVIQDYDNQHEODRPTSPAPSRFFSVLR 466
 CC 290 PRPVSYLK-----GSSGGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPV 333
 CC 467 ANDVLMILSLTAAYDOSTYSGSTGPPYV--SDSVTLVNVATGAQAVASIDMTKV 519
 CC
 CC RESULT 11
 CC SRSC_ARATH
 CC ID SRSC_ARATH STANDARD; PRT; 564 AA.
 CC AC P37107; 082570;

DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Signal recognition particle 54 kDa protein, chloroplast precursor
 DE (SRP54) (54 chloroplast protein) (54CP) (FFC).
 GN PFC OR AFG03940 OR F8F6_150.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eucroids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX STRAIN=cv. Columbia;
 RX MEDLINE=94012817; PubMed=8408079;
 RA Franklin A.E., Hoffman N.E.;
 RT "Characterization of a chloroplast homologue of the 54-kDa subunit of
 the signal recognition particle";
 RL J. Biol. Chem. 268:22175-22180 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;
 RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc,
 encoding the 54 kDa subunit of chloroplast signal recognition
 particle";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RC MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneo T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasaki S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
 RA Nakazaki N., Natsu K., Okumura S., Shino S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston J.,
 RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes J., Haekens B., Jamar E., Latreille P.,
 RA Du R., Edwards J., Fryman J., Haekens B., Jamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
 RA Wagner-McPherson C., Wolian A., Yeakum M., Bell M., Dedila N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
 RA Mattiensen R., McComb W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wambut R., Duesterhoeft A., Strikema W., Pohl T.,
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Langham S.-A., McCullagh B., Robben J., Grymptre B., Zimmermann W.,
 RA Ramsparger U., Wedler H., Balke K., Wedler E., Peters S.,
 RA van Steveren M., Dikse W., Mooijman P., Klein Lankhorst R.,
 RA Weitzenecker T., Bothe G., Rose M., Hauf J., Bernerstorfer S., Hempel S.,
 RA Feldpausch M., Lamberth S., Villarroel R., Gelen J., Ardiles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.,
 RA "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana";
 RT Nature 408:823-826 (2000).
 RL Nature 408:823-826 (2000).
 CC -1- FUNCTION: May target chloroplast proteins to either the thylakoid
 or envelope membranes.
 CC -1- SUBCELLULAR LOCATION: Chloroplast stroma.
 CC -1- TISSUE SPECIFICITY: Most abundant in green shoot tissue and
 lower levels seen in the roots and etiolated buds.
 CC -1- SIMILARITY: Belongs to the GTP-binding SRP family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL, Z21970; CAA79981.1; -

DR EMBL, AF092168; AAC64139.1; -
 DR EMBL, AL162873; CAB85514.1; -
 DR PIR, S36637; S36637.
 DR HSSP, C07347; IFFH.
 DR InterPro: IPR003593; AAA ATPase.
 DR InterPro: IPR000897; SRP54.
 DR InterPro: IPR004125; SRP54_SBP.
 DR InterPro: IPR004780; SRP_SBP.
 DR Pfam: PF00448; SRP54_1.
 DR Pfam: PF02881; SRP54_N; 1.
 DR Pfam: PF02978; SRP_SBP; 1.
 DR ProDom: PD000819; SRP54_1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRFam: TIGR00959; ffb; 1.
 DR PROSITE, PS00300; SRP54_1.
 KW Signal recognition particle, GTP-binding, RNA-binding, Chloroplast;
 KW Transist peptide.
 FT TRANSIT 1
 FT CHAIN 76
 FT DOMAIN 76 370
 FT DOMAIN 371 564
 FT NP_BIND 183 190
 FT NP_BIND 265 269
 FT NP_BIND 323 326
 FT CONFLICT 76 76
 SQ SEQUENCE 564 AA; 61232 MW; 423F7285FB906384 CRC64;
 Query Match 5.7%; Score 101; DB 1; Length 564;
 Best Local Similarity 26.1%; Freq. No. 0.6;
 Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;
 QY 15 VPEIET-----PDITLLAIGPLMVLQAGI-----TKVFFVPAQGLIRACMLVR 61
 DB 154 VHDVLYKLMGGEVSELOKSGPVTYLLAGLGQVGTCTVCKALCYLKKQ--KSCWLI- 210
 QY 62 KAAGGVYQVAPFKIALI---TGYVYDHLPLQ--DVAHAGLADLVAAPVVFSSMEV 116
 DB 211 --AGVYPPAIDVLVIGGVGVYTAGDVPRADIAKGLKFAK-----NNVDV 261
 QY 117 KITTGADTAACGDISGLPVSARSGREI-----LGPADNFEQGWRLAPITAVSQOT 171
 DB 262 VIM---DTAGRLQIDKGMDELKDKFLEPTEVLVLDMTQ--EAAALVTTFVVEI 315
 QY 172 RGLGCIITSLTGRDKNQGEVEQVYS 198
 DB 316 -GITGAILTKLDGDSRGGAALSVKVS 341
 RESULT 12
 ID DP02 MOUSE STANDARD; PRT; 600 AA.
 AC P33611;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
 GN POLA2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 265-285 AND 394-403.
 RX MEDLINE=93216788; PubMed=8463324;
 RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
 RA Hanaoka F.;
 RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA
 polymerase alpha-primase complex and their gene expression during
 cell proliferation and the cell cycle";
 RL J. Biol. Chem. 265:8111-8122 (1993).
 CC -1- FUNCTION: May play an essential role at the early stage of
 chromosomal DNA replication by coupling the polymerase

CC alpha/primase complex to the cellular replication machinery (By
CC similarity).
CC -1- SUBUNIT: DNA polymerase alpha-primase is a four subunit enzyme
CC (subunits A, B, C and D), which is assembled throughout the cell
CC cycle. The largest subunit (subunit A) has DNA polymerase
CC activity, the two smallest subunits (subunits C and D) have DNA
CC primase activity. Subunit B binds to subunit A.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M
CC PHASE (BY SIMILARITY).
CC -1- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D13546; BAA02746.1; -
CC PIR: B46642; B46642.
CC MGI: 99690; Polaz.
CC InterPro: IPR007200; DNA_pol_alpha_B.
CC Pfam: PF04058; DNA_pol_alpha_B; 1.
CC KEGG: K01017; POLY-GLU.
CC FT DOMAIN 101 107
CC FT 115 157 PRO/SER/THR-RICH (HYDROPHILIC).
CC SQ SEQUENCE 600 AA; 66267 MW; 79F94BEEF33FEEC CRC64;

Query Match 5.7%; Score 101; DB 1; Length 600;
Best Local Similarity 24.8%; Pred. No. 0.65; Indels 62; Gaps 12;
Matches 55; Conservative 34; Mismatches 71;

QY 53 LIRACMVRKAGHYVOM-APMKAALT- -GTYYVDHL- -TPLDMA 95
Db 27 LAELCVLYRQEDMVBELIFCTSAKTCITLVDLNSFEVYANKLSAKMSASDSC 86
QY 96 HAGRIDLAIVEPIVSDMEYKITTQADTACDI- -ISGLP- -VSARRGRET 145
Db 87 HAGRID- -VSIOELIEAEEHEETLLSSYTPSKPLKRVSTPEPTPKASVAARSRC- 144
QY 146 LLGPAIDFEGGQWRLAPITRYSQOTRGILGCIITSLTGRDKXQVEVEQVYVATQSF 205
Db 145 LLSBS- - - - -FSPSATPSOK- - - - -YTSSTNR- - - - -GEVTTGSAQ- - - 178
QY 206 ATCVNGVCWTFVHGAGSKTL- -AGPKGPIQMTYVNDQDLVG 245
Db 179 - - - - -GLWSGSGSGSVSLKVVGDPEPLTGSYVAMFQQLMG 215

RESULT 13
CYAA LEIDO
ID CYAA LEIDO STANDARD; PRT; 1380 AA.
AC Q27675;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Receptor-type adenylylate cyclase A (EC 4.6.1.1) (ATP pyrophosphate-
DE lyase) (Adenylyl cyclase).
GN RAC-A.
OS Leishmania donovani.
OC Eukaryota; Eukleozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
ON NCBI_TaxID=5661;
RX MEDLINE=5340554; Pubmed=7615561;
RA Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M.,
RT "A family of putative receptor-adenylylate cyclases from Leishmania
RT donovani";
RL J. Biol. Chem. 270:17551-17558 (1995).
CC -1- FUNCTION: Could act as a receptor for a unknown ligand.

CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote)
CC but not in the mammalian host stage of the parasite life cycle.
CC -1- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U17042; AAA7498.1; -
CC PIR: T18309; T18309.
CC InterPro: IPR001054; G_cyclase.
CC Pfam: PF00211; guanylate_cyc; 1.
CC SMART: SM00044; CYCC; 1.
CC DR PROSITE: PS50125; GUANYLATE CYCLASES 2; 1.
CC KW Metal-binding; Magnesium; Transmembrane; Receptor; Glycoprotein;
CC Metal-binding; Magnesium.
CC FT DOMAIN 1 34
CC FT TRANSMEM 35 55
CC FT DOMAIN 56 891
CC FT TRANSMEM 892 912
CC FT DOMAIN 913 1380
CC FT METAL 938 938
CC FT METAL 981 981
CC FT CARBOHYD 422 422
CC FT CARBOHYD 478 478
CC FT CARBOHYD 497 497
CC FT CARBOHYD 567 567
CC SQ SEQUENCE 1380 AA; 151692 MW; 6B2D5F7D3C1107A0 CRC64;

Query Match 5.3%; Score 93.5; DB 1; Length 1380;
Best Local Similarity 21.0%; Pred. No. 7.8;
Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;

QY 3 GGRDAITLTCAVHP- -ELIFDITK- - - - -LLAFGLM- - - - -V 36
Db 101 GGRPIKIL- - - - -HPDDQDNLYDAEVLHSLAQEKLLAVGLGPELDRLTLAALSNDV 155
QY 37 LQAGITKVP- - - - -YFPAQGLIRACMVRKAGHYV- - - - -QMAENK 75
Db 156 VQSGELILAPFTGSSGVRTMSDVYFTFAEPVVELKVLN- - - - -HIVRLRAREVAFVR 210
QY 76 LAAITG- - - - -PYVDHLTFLQDMHAGIRDLA- - - - -VERVITSDVEYK- - - - - 117
Db 211 - - - - -LFGMHFGGEELTYVDLTLSL- - - - -LRDPAVLYTYVSSSVVEDEAFDAMAD 260
QY 118 - - - - -ITWGADTAAACGDIISGLPVASARRGRIILG- - - - -PADNFEQ 156
Db 261 TNPQVITVAAPVQGVITFLEKVLDPRTSSAYVISCMIQKVPADVYKRLLSASIRPQ 320
QY 157 GKRLLAPITAYSCQTRGLIGCIITSLTGRDKQVGE- - - - -VQVSTATQSLATCVNGV 212
Db 321 DGRITIASAT- - - - -SPVSGEGLKXMEVLVAKQMSNYIE- - - - -NSG 356
QY 213 CWTVEHGAGSKTLAPK- - - - -GPITQMTYVNDQDLVGMQAPGASMTPTCTGSSDLYLV 268
Db 357 SDIYVPPDDSTLTGRKARSEAPLSRKT- - - - -FQAHSTIKMALGLSTLVQ 412
QY 269 TRHADVTPIYRRRGRSGSLSPRY- - - - -SYLKGSGGPLL- - - - -PSGHAVGITRAAVCTRGV 324
Db 413 TLEQTDWIVNRYKAGLGNQRRFVIGDYLVDYGGP- - - - -CEPLAQFLG- - - - -ASCCNOGG 468
QY 325 AKAV- - - - - -DFIPVSM 337
Db 469 HSSITLVQNASWDIVPDSFR 490

```

RESULT 14
TOOLB_CHLITE STANDARD; PRT; 434 AA.
ID TOOLB_CHLITE
AC Q8KE00;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE TolB protein precursor.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RE SEQUENCE FROM N.A.
RC SRAIN-TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Niernan W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vanatrevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium."
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
CC -1- FUNCTION: Involved in the comb-independent uptake of proteins (by
CC similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: Belongs to the tolB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB012837; AAM71875.1; -
CC TIGR; CT0636; -
CC HAMAP; MF 00671; -; 1.
CC Transport; protein transport; Periplasmic; Signal; Complete proteome.
CC SIGNAL
CC FT 28 434 TolB protein.
CC CHAIN
CC SEQUENCE 434 AA; 47275 MW; 1FA347B99C7A0F99 CRC64;
CC
Query Match 5.2%; Score 93; DB 1; Length 434;
Best Local Similarity 21.2%; Pred. No. 2.1;
Matches 69; Conservative 47; Mismatches 139; Indels 70; Gaps 14;
QY 31 FGLPVLQAGITKQPYEYFAQGLIRACKLVKKAAGHYVQMAFKALALGTYYVYDHLTP 90
DB 8 FACLCIMLFMLFVFETLRAEE-VGEYIAIRK-EGASRIAVVDKTSADGKQ----- 58
QY 91 LQDMAHA-----GLRLAVNEPVIFSDMEVKITWQADPACGDIISGLPVS 138
DB 59 -REWARSIDVTINKGLDFTGSLFNILPA-PLINANGNGGNGNFASISVSGDIAGGSVT 115
QY 139 ARRGSEIL-----LGPADNFEQGMRLAPITAVYQQRGLIGCITTS 181
DB 116 KRSGRPVLEMHVYDSSGKSLARTYGESQLRAIGRFCADLVLELLTGKSVFGRIVF 175
QY 182 LTGRPKNQ-----VEGEVYVSTATQSFATCV-----NGVCYTVHGAAGSKTLAPKG 230
DB 176 VANTRGNEIYICDFDGENVQLTNSRISILTPAVSPGITYANTDY--TSGRKNLYIKN 233
QY 231 PITQWYTVNPDOLV---GWOAPGARSMTPCTC--GSSDLYLTVTRADVIPVRRGDSR 284
DB 234 IATGAKVSVNHGVCISPAWR--PGNTLTLLTTSYEGQDYLXL--RADGTVERRLTKGG 289
QY 285 GSLISP-----RPVSYIKGSSGSP 303

```

```

DB 290 GIDVSTFSPDGSQMAFVSTRGSP 314
RESULT 15
TREE HORSE STANDARD; PRT; 706 AA.
ID TREE_HORSE
AC P27425;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal
DE binding globulin).
DE TP.
GN Equus caballus (Horse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RE SEQUENCE FROM N.A.
RC MEDLINE=93277858; PubMed=8504171;
RA Carpenter M.A., Broad T.E.;
RT "The cDNA sequence of horse transferrin."
RL Biochim. Biophys. Acta 1173:230-232(1993).
RN [2]
RE SEQUENCE FROM N.A.
RC TISSUE-Extraembryonic tissue;
RA McDowell K.J., Adams M.H., Baker C.B.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Transferrins are iron binding transport proteins which
CC can bind two atoms of ferric iron in association with the binding
CC of an anion, usually bicarbonate. It is responsible for the
CC transport of iron from sites of absorption and heme degradation to
CC those of storage and utilization. Serum transferrin may also have
CC a further role in stimulating cell proliferation.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -1- DOMAIN: Composed of two homologous domains.
CC -1- SIMILARITY: Belongs to the transferrin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M69020; AAA0958.1; -
CC EMBL; U21127; AAA63684.1; -
CC PIR; S33761; S33761.
CC HSSP; P02787; 1A85.
CC InterPro: IPR001156; Transferrin.
CC Pfam: PF00405; transferrin.
CC SMART; SM00094; TR_FER; 2.
CC PROSITE; PS00205; TRANSFERRIN_1; 2.
CC PROSITE; PS00206; TRANSFERRIN_2; 2.
CC PROSITE; PS00207; TRANSFERRIN_3; 2.
CC Transport; iron transport; Glycoprotein; Metal-binding; Repeat;
CC SIGNAL
CC FT 1 19 BY SIMILARITY.
CC CHAIN 20 706 SEROTRANSFERRIN.
CC REPEAT 20 357 1.
CC FT REPEAT 358 706 2.
CC FT DISULFID 26 64 BY SIMILARITY.
CC FT DISULFID 36 55 BY SIMILARITY.
CC FT DISULFID 134 235 BY SIMILARITY.
CC FT DISULFID 174 190 BY SIMILARITY.
CC FT DISULFID 177 198 BY SIMILARITY.
CC FT DISULFID 187 200 BY SIMILARITY.
CC FT DISULFID 248 262 BY SIMILARITY.

```



```
FT DISULFID 360 623 BY SIMILARITY.
FT DISULFID 366 398 BY SIMILARITY.
FT DISULFID 376 389 BY SIMILARITY.
FT DISULFID 423 701 BY SIMILARITY.
FT DISULFID 441 664 BY SIMILARITY.
FT DISULFID 474 550 BY SIMILARITY.
FT DISULFID 498 692 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.
FT DISULFID 519 533 BY SIMILARITY.
FT DISULFID 590 604 BY SIMILARITY.
FT DISULFID 642 647 BY SIMILARITY.
FT METAL 79 79 BY SIMILARITY.
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 209 209 IRON 1 (BY SIMILARITY).
FT METAL 270 270 IRON 1 (BY SIMILARITY).
FT METAL 413 413 IRON 2 (BY SIMILARITY).
FT METAL 449 449 IRON 2 (BY SIMILARITY).
FT METAL 544 544 IRON 2 (BY SIMILARITY).
FT METAL 612 612 IRON 2 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 706 AA; 78094 MM; 1A0FA566C0409D8A CRC64;
```

Query Match 5.2%; Score 92.5; DB 1; Length 706;

Best Local Similarity 21.5%; Pred. No. 4.2; Indels 81; Gaps 17;

Matches 64; Conservative 43; Mismatches 109;

```
QY 71 MAFKMLALGTGVY---YDHLTPLODWAHAGRLDVAVEVIFSDMEVKIITWGA----123
Db 321 LGFLRIPAMDWTWYLGVEYVT-----AIRNLREDIREVPKD-ECKKVKWCAIGHH 371
QY 124 DTAACGD--ITSGLPVBARBR-----EILGPADNPEGQGWRL-----LAPITAY 167
Db 372 EKVKCDMSVNSGNIGCESAOSTEDCIATKIVKGEADAMSIDGGFIYIAGKCGIVPVLAE 431
QY 168 SOOTRGLIGCIITSLTGSDKNQVEGVVSTATOSFIATCVNGVCTVPHGAGSKTIAG 227
Db 432 NYETRRSGSACVDTPEEGH-----AVAVKSSSDPDLT-----W-----NSLKG 470
QY 228 PKGPTQMTYNNVDOLVGMQAPPGARSMPTCTGSSDLYLVTRHADVIVPERRGDSRGL 287
Db 471 KK---SCHTGVDR-TAGWNI PMGL-----LYSEIKHCEFDKFFREGCAPGYR 513
QY 288 LSPRPVSLKSSGSGP--LLC-PSGHA-----VGIFRAVCTRGVAKAVDPIPVESME 337
Db 514 RNSTLCNLCIGSASGPGRECEPNNHRYGYTGAFRLCLVEKGDVA---FVKHQYVB 566
```

Search completed: May 6, 2004, 09:31:54

Job time : 8.119 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 31.8931 Seconds
(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-14
Perfect score: 1778
Sequence: 1 VFGGRDAIILTCVAHPBLI.....RGVAKAVDFIPVESMETWR 341

Scoring table: BIOSUM62
Gapcp 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTRMBL_25:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_yerebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1729	97.2	3010	12 Q9DTB6	Q9dte6 hepatitis c
2	1726	97.1	3010	12 Q9J3P9	Q9j3i9 hepatitis c
3	1721	96.8	3010	12 Q9D7D6	Q9dte6 hepatitis c
4	1719	96.7	3010	12 Q9J3H7	Q9j3h7 hepatitis c
5	1719	96.7	3010	12 P90191	P90191 hepatitis c
6	1718	96.6	3010	12 Q9J3H5	Q9j3h5 hepatitis c
7	1715	96.5	3008	12 Q9J3F4	Q9j3f4 hepatitis c
8	1715	96.5	3010	12 Q9J3H3	Q9j3h3 hepatitis c
9	1715	96.5	3010	12 Q9D7D7	Q9dte7 hepatitis c
10	1715	96.5	3010	12 Q807P3	Q807p3 hepatitis c
11	1714	96.4	3010	12 P88803	P88803 hepatitis c
12	1711	96.2	1186	12 Q81755	Q81755 hepatitis c
13	1711	96.2	2284	12 Q81817	Q81817 hepatitis c
14	1711	96.2	3010	12 Q68788	Q68788 hepatitis c
15	1711	96.2	3010	12 P89366	P89366 hepatitis c
16	1710	96.2	3010	12 Q99A02	Q99a02 hepatitis c

17	1710	96.2	3010	12 Q9QIX6	Q9qix6 hepatitis c
18	1710	96.2	3010	12 Q9QIX5	Q9qix5 hepatitis c
19	1710	96.2	3010	12 Q68826	Q68826 hepatitis c
20	1710	96.2	3010	12 Q9J3G6	Q9j3g6 hepatitis c
21	1710	96.2	3010	12 Q9DTE4	Q9dte4 hepatitis c
22	1709	96.1	3010	12 Q9J3H9	Q9j3h9 hepatitis c
23	1708	96.1	3010	12 Q9J3I0	Q9j3i0 hepatitis c
24	1708	96.1	3010	12 Q09796	Q09796 hepatitis c
25	1708	96.1	3011	12 Q9DTE3	Q9dte3 hepatitis c
26	1707	96.0	3010	12 Q9QIY3	Q9qi y3 hepatitis c
27	1705	95.9	3010	12 Q9QIY5	Q9qi y5 hepatitis c
28	1705	95.9	3014	12 Q9DTE0	Q9dte0 hepatitis c
29	1704	95.8	3010	12 Q9DTE0	Q9dte0 hepatitis c
30	1703	95.8	3010	12 Q02829	Q02829 hepatitis c
31	1702	95.7	3010	12 Q9J3H0	Q9j3h0 hepatitis c
32	1702	95.7	3010	12 Q81541	Q81541 hepatitis c
33	1702	95.7	3010	12 Q9WMX2	Q9wm x2 hepatitis c
34	1701	95.7	3010	12 Q81760	Q81760 hepatitis c
35	1701	95.7	3010	12 Q9QIX8	Q9qi x8 hepatitis c
36	1701	95.7	3010	12 Q9QIX7	Q9qi x7 hepatitis c
37	1701	95.7	3010	12 Q8QRL8	Q8qrl8 hepatitis c
38	1701	95.7	3010	12 Q9J3H6	Q9j3h6 hepatitis c
39	1699	95.6	3010	12 Q9J3I1	Q9j3i1 hepatitis c
40	1698	95.5	3010	12 Q9J3H2	Q9j3h2 hepatitis c
41	1698	95.5	3010	12 Q9DTE5	Q9dte5 hepatitis c
42	1698	95.5	3010	12 Q9DTE7	Q9dte7 hepatitis c
43	1698	95.5	3010	12 Q9QIY4	Q9qi y4 hepatitis c
44	1697	95.4	3010	12 Q91A00	Q91a00 hepatitis c
45	1697	95.4	3010	12 Q9J3G3	Q9j3g3 hepatitis c

ALIGNMENTS

RESULT 1
Q9DTE6 PRELIMINARY; PRT; 3010 AA.
ID Q9DTE6
AC Q9DTE6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV142;
RA Takahashi K., Iwata K., Matsumoto H., Nakao K.,
RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hikioka M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
with hepatocellular carcinoma: the 'progression score' revisited.";
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; AB049091; BAB1804.1; -.
CC PIR; A61196; A61196.
DR PIR; PS0329; PS0329.
DR HSSP; P28663; IUXP.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019828; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006118; P:electron transport; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser lyase.
 DR InterPro: IPR000345; Cys Ser lyase.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR004109; Peptidase C2.
 DR InterPro: IPR007095; RNA pol D5 PS.
 DR InterPro: IPR007094; RNA pol PS1r.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; Viral RdRp; 1.
 DR ProDom: PD18662; HCV NS1; 1.
 DR SMART: SM00487; DEAD; 1.
 DR SMART: SM00490; HELIC; 1.
 DR PROSITE: PS00190; CYTOCHROME C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolyase; Nonstructural protein; Polypeptide;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 97.2%; Score 1729; DB 12; Length 3010;
 Best Local Similarity 95.9%; Pred. No. 1,36-138;
 Matches 327; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 VRGRDAIILITCAVHELIIFDITKILAIIFGPIWVZAGITVVFVRAQGLIRACMLV 60
 DB 866 VRGRDAIILITCAVHELIIFDITKILAIIFGPIWVZAGITVVFVRAQGLIRACMLV 925

QY 61 RKAAGHYVQMAFKALITGTYVDLITPLQDMAHGLDLAVAYEVYFSMEYKILIT 120
 DB 926 RKAAGHYVQMAFKALITGTYVDLITPLQDMAHGLDLAVAYEVYFSMEYKILIT 985

QY 121 WGAADTAACGDIISGLPVSARREIILGPADNFGQGMRIADITVYSCOTRGLICIT 180
 DB 986 WGAADTAACGDIISGLPVSARREIILGPADNFGQGMRIADITVYSCOTRGLICIT 1045

QY 181 SLTRDKNQVGEVQVSTATQSFATVGVCTVFGAGSKTLGPKPITQMTYND 240
 DB 1046 SLTRDKNQVGEVQVSTATQSFATVGVCTVFGAGSKTLGPKPITQMTYND 1105

QY 241 QDLVGMQAPPGARSMCTCGSDLYLVTHADVIFPRRGDSRGSLSPRPSYKGGSS 300
 DB 1106 QDLVGMQAPPGARSLTCTCGSSDLYLVTHADVIFPRRGDSRGSLSPRPSYKGGSS 1165

QY 301 GGPILCPGSHAVGIFRAVCTRGVAKAVDIPVESMETTNR 341
 DB 1166 GGPILCPGSHAVGIFRAVCTRGVAKAVDIPVESMETTNR 1206

RESULT 2

Q9J3F9 ID Q9J3F9 PRELIMINARY; PRT; 3010 AA.
 AC Q9J3F9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBT_Taxid=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD3;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease progression."
 RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL: AF207774; AAF65964.1; -.
 DR PIR: A61196; A61196.
 DR PIR: P00246; P00246.
 DR PIR: P03329; P03329.
 DR HSSP: P27958; 1HE1.
 DR MEROPS: S29.001; -.
 DR GO:0016021; C:integral to membrane; IEA.
 DR GO:0019028; C:viral envelope; IEA.
 DR GO:0019031; C:viral envelope; IEA.
 DR GO:0005524; F:ATP binding; IEA.
 DR GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO:0005489; F:electron transporter activity; IEA.
 DR GO:0003723; F:RNA binding; IEA.
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO:0008235; F:serine-type peptidase activity; IEA.
 DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006118; P:electron transport; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser lyase.
 DR InterPro: IPR000345; Cys Ser lyase.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR004109; Peptidase C2.
 DR InterPro: IPR007095; RNA pol D5 PS.
 DR InterPro: IPR007094; RNA pol PS1r.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; Viral RdRp; 1.
 DR ProDom: PD18662; HCV NS1; 1.

DR SMART, SM00487, DEXdc, 1.
 DR PROSITE; PS00190; CYTOCHROME C, 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polyprotein; RNA-directed RNA polymerase; Glycosylase; Transferrase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C9DB93E60C7 CRC64;

Query Match 97.1%; Score 1726; DB 12; Length 3010;
 Best Local Similarity 95.6%; Pred. No. 2.4e-138;
 Matches 326; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAFGLPMTLQAGITKPYFVRAQGLIRACMLV 60
 DB 866 VRGGRDAIILITCAVHPELIFDITKLLAFGLPMTLQAGITKPYFVRAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLPLOMAHAGRLDAVAEPIFSMEVKIIT 120
 DB 926 RKAAGHYVQMAFMKLAALGTYYVDHLPLOMAHAGRLDAVAEPIFSMEVKIIT 985
 QY 121 WGAUTAACGDIISGLPVASRGREIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 180
 DB 986 WGAUTAACGDIISGLPVASRGREIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 1045
 QY 181 SLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKSPITQMTYND 240
 DB 1046 SLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKSPITQMTYND 1105
 QY 241 QDLVGMQAPPGASMTCTCGSSDLVLTNRADVIPRRRGDSRGSLLSPRVSYLKGS 300
 DB 1106 QDLVGMQAPPGASMTCTCGSSDLVLTNRADVIPRRRGDSRGSLLSPRVSYLKGS 1165
 QY 301 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 341
 DB 1166 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 3
 Q9DPT6 PRELIMINARY; PRT; 3010 AA.
 ID Q9DPT6
 AC Q9DPT6
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Genome Polyprotein.
 OS Hepatitis C virus.
 OC Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CX NCBI_taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV221;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 3A Hatahara T., Ohno Y., Kanai K., Maruo H., Baba K., Hijioka M.,
 RA Mishiro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 with hepatocellular carcinoma: the progression score, revisited.";
 RL Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA (BY SIMILARITY).
 CC EMBL; AB049101; BAB18814.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; PS0329; PS0329.
 DR HSSP; P26663; 1uxp.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0005489; F:electron transporter activity; IEA.
 DR GO; GO:0016787; F:helicase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006118; F:electron transport; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; F:transcription; IEA.
 DR GO; GO:0019079; F:viral genome replication; IEA.
 DR GO; GO:0019087; F:viral transformation; IEA.
 DR GO; GO:0009003; Cys_ser_tyrosin.
 DR InterPro; IPR000345; Cys_ser_tyrosin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV capsid.
 DR InterPro; IPR002521; HCV core.
 DR InterPro; IPR002519; HCV env.
 DR InterPro; IPR002511; HCV NS1.
 DR InterPro; IPR002518; HCV NS2.
 DR InterPro; IPR000745; HCV NS4a.
 DR InterPro; IPR001490; HCV NS4b.
 DR InterPro; IPR002868; HCV NS5a.
 DR InterPro; IPR002166; HCV RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV capsid; 1.
 DR Pfam; PF01542; HCV core; 1.
 DR Pfam; PF01539; HCV env; 1.
 DR Pfam; PF01560; HCV NS1; 1.
 DR Pfam; PF01568; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF02907; HCV NS4a; 1.
 DR Pfam; PF01006; HCV NS4b; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRp; 1.
 DR ProDom; PD186062; HCV NS1; 1.
 DR SMART; SM00487; DEXdc; 1.
 DR PROSITE; PS00190; CYTOCHROME C, 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolyase; Nonstructural protein; Polyprotein;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327108 MW; DE182D810EF78E84 CRC64;

Query Match 96.8%; Score 1721; DB 12; Length 3010;
 Best Local Similarity 95.9%; Pred. No. 6.5e-138;
 Matches 327; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAFGLPMTLQAGITKPYFVRAQGLIRACMLV 60
 DB 866 VRGGRDAIILITCAVHPELIFDITKLLAFGLPMTLQAGITKPYFVRAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAFMKLAALGTYYVDHLPLOMAHAGRLDAVAEPIFSMEVKIIT 120
 DB 926 RKAAGHYVQMAFMKLAALGTYYVDHLPLOMAHAGRLDAVAEPIFSMEVKIIT 985
 QY 121 WGAUTAACGDIISGLPVASRGREIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 180
 DB 986 WGAUTAACGDIISGLPVASRGREIILGPADNFEQGMRLAPITAYSOQTRGLLCIIT 1045
 QY 181 SLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKSPITQMTYND 240
 DB 1046 SLTGRDNQVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAPKSPITQMTYND 1105
 QY 241 QDLVGMQAPPGASMTCTCGSSDLVLTNRADVIPRRRGDSRGSLLSPRVSYLKGS 300
 DB 1106 QDLVGMQAPPGASMTCTCGSSDLVLTNRADVIPRRRGDSRGSLLSPRVSYLKGS 1165
 QY 301 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 341
 DB 1166 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 4

Q9J3H7 PRELIMINARY; PRT; 3010 AA.

AC Q9J3H7;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI_Taxid=11103;

RE SEQUENCE FROM N.A.

RC STRAIN=MD5; Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.,

RA Nagayama K., Characteristics of hepatitis C viral genome associated with disease

RT progression."

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

PROTEIN C AND RNA (BY SIMILARITY).

CC EMBL; AF20756; AAF65946.1; -

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; P00804; P00804.

DR PIR; PS0329; PS0329.

DR HSSP; P26663; IUXP.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F:electron transporter activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008262; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006118; F:electron transport; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P:transcription; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR GO; GO:0019087; P:viral transformation; IEA.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR000345; CysC_heme_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002531; HCV_NS1.

DR InterPro; IPR002518; HCV_NS2.

DR InterPro; IPR000745; HCV_NS4a.

DR InterPro; IPR001490; HCV_NS4b.

DR InterPro; IPR002686; HCV_NS5a.

DR InterPro; IPR002168; HCV_NS5a.

DR InterPro; IPR004109; Peptidase_C29.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV_NS1; 1.

DR Pfam; PF01538; HCV_NS2; 1.

DR Pfam; PF02907; HCV_NS3; 1.

DR Pfam; PF01006; HCV_NS4a; 1.

DR Pfam; PF01001; HCV_NS4b; 1.

DR Pfam; PF01506; HCV_NS5a; 1.

DR Pfam; PF00998; Viral_Rdrp; 1.

DR ProDom; PD16062; HCV_NS1; 1.

DR SMART; SM00487; DEXDc; 1.

DR PROSITE; PS00190; CYTOCHROME_C; 1.

KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

KW Polyprotein; RNA-directed RNA polymerase; transferase; transmembrane.

SC SEQUENCE 3010 AA; 327365 MW; D8653F7317FFA106 CRC64;

Query Match 96.7%; Score 1719; DB 12; Length 3010;

Best Local Similarity 95.0%; Pred. No. 9.6e-138;

Matches 324; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 1 VGGSDAIIILLTCAVHPELLIDITKLLAIIGPLMTLOAGITKQYFVPAQGLIRACMLV 60

DB 866 VGGSDAIIILLTCAVHPELLIDITKLLAIIGPLMTLOAGITKQYFVPAQGLIRACMLV 925

QY 61 RKAGGHVQVQAFMLTALITGVYVDHLTPQDMVHAGLADLAVERVPSDMEVKIIT 120

DB 926 RKAGGHVQVQAFMLTALITGVYVDHLTPQDMVHAGLADLAVERVPSDMEVKIIT 985

QY 121 WQADTAACGDIISGLPVASRGRREILGPADNFEQGRLLAPITVYSQOTRGLIGCIIT 180

DB 986 WQADTAACGDIISGLPVASRGRREILGPADNFEQGRLLAPITVYSQOTRGLIGCIIT 1045

QY 181 SLTGRDNQVGEVQVVSSTATOSFLATGVNVCMTVFGAGSKTLAGKPIITQMTYND 240

DB 1046 SLTGRDNQVGEVQVVSSTATOSFLATGVNVCMTVFGAGSKTLAGKPIITQMTYND 1105

QY 241 QDLVGMQAPFGARSMTPTCCSSDLVYVTRHADVIPVRRDSSGSLSPRVSYLKGS 300

DB 1106 QDLVGMQAPFGARSMTPTCCSSDLVYVTRHADVIPVRRDSSGSLSPRVSYLKGS 1165

QY 301 GGPLICPSGHAVGIFRAVCTRGVAKAVDFIPVSMETMTM 341

DB 1166 GGPLICPSGHAVGIFRAVCTRGVAKAVDFIPVSMETMTM 1206

RESULT 5

P90191 PRELIMINARY; PRT; 3010 AA.

AC P90191;

DT 01-MAY-1997 (TREMBLrel. 03, Created)

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI_Taxid=11103;

RE SEQUENCE FROM N.A.

RC STRAIN=HCV-1b.

RA Enomoto N.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=HCV-1b.

RX MEDLINE=95340824; PubMed=7542279;

RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,

RA Yamamoto C., Izumi N., Marumo F., Sato C.,

RT "Comparison of full-length sequences of interferon-sensitive and

RT resistant hepatitis C virus 1b."

RT J. Clin. Invest. 96:224-230 (1995).

-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

PROTEIN C AND RNA (BY SIMILARITY).

CC EMBL; D50482; BAA09073.1; -

DR PIR; A61196; A61196.

DR PIR; P00254; P00254.

DR PIR; P00804; P00804.

DR PIR; PS0329; PS0329.

DR PDB; 1DXW; 12-JAN-01.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO:0003723; F:RNA binding; IEA.
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO:0008286; F:serine-type peptidase activity; IEA.
 DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0001674; F:transferase activity; IEA.
 DR GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO:0006350; F:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR001490; HCV_NS4a.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 KW Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191 CORE PROTEIN.
 FT CHAIN 192 383 E1.
 FT CHAIN 384 809 E2.
 FT CHAIN 810 1026 NS2.
 FT CHAIN 1027 1657 NS3.
 FT CHAIN 1658 1711 NS4a.
 FT CHAIN 1712 1972 NS4b.
 FT CHAIN 1973 2419 NS5a.
 FT CHAIN 2420 3010 NS5b.
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match 96.7%; Score 1719; DB 12; Length 3010;
 Best Local Similarity 95.3%; Pred. No. 9,66-138;
 Matches 325; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 1 VRGGDAIILITCAVHEPLIDITKLLAIFGLMTLOAGITKPYVRAQGLIRACMLV 60
 DB 866 VRGRDAIILITCAVHSELVEITKLLAIFGLMTVLOAGITKPYVRAQGLIRACMLV 925
 QY 61 RKAAGHVVQVAFKMLALGTGYVYDHLTPQDMVAHAGLRDLAAVAVPVEFSDMEVKIT 120
 DB 926 RKVAGHVVQVAFKMLALGTGYVYDHLTPQDMVAHAGLRDLAAVAVPVEFSDMEVKIT 985
 QY 121 WGADTAACGDIISGLPVASRGREIILGPADNFEQGQRLIAPITAVSQOTRGLIGCIIIT 180
 DB 986 WGADTAACGDIISGLPVASRGREIILGPADNFEQGQRLIAPITAVSQOTRGLIGCIIIT 1045
 QY 181 SLTRGDKVQGEVGVVSTATOSFLATGVNVCMTVEFGAGSKTLGKGIITOMYTNVD 240
 DB 1046 SLTRGDKVQGEVGVVSTATOSFLATGVNVCMTVEFGAGSKTLGKGIITOMYTNVD 1105
 QY 241 QDLVGMQAPPGARSMTPTCGSSSDLYLVTRHADVI PVRRGDSRGSLLSPRPVSYLKGS 300

DB 1106 QDLVGMQAPPGARSMTPTCGSSSDLYLVTRHADVI PVRRGDSRGSLLSPRPVSYLKGS 1165
 QY 301 GGPLCPSGHANGIRFANCTGKVAKADVPIVSEMETTMR 341
 DB 1166 GGPLCPSGHANGIRFANCTGKVAKADVPIVSEMETTMR 1206
 RESULT 6
 ID Q9J3H5 PRELIMINARY; PRT; 3010 AA.
 AC Q9J3H5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN-MD17;
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease progression."
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; AF207158; AA65948.1; -
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P00254; P00254.
 DR PIR; P50329; P50329.
 DR HSSP; P27938; IHEI.
 DR GO:0016021; C:Integral to membrane; IEA.
 DR GO:0019028; C:viral capsid; IEA.
 DR GO:0019031; C:viral envelope; IEA.
 DR GO:0005524; F:ATP binding; IEA.
 DR GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO:0005489; F:electron transporter activity; IEA.
 DR GO:0016787; F:hydrolase activity; IEA.
 DR GO:0003723; F:RNA binding; IEA.
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006118; F:electron transport; IEA.
 DR GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019079; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_trypsin.
 DR InterPro: IPR000345; Cys_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.
 DR Pfam; PF02907; HCV NS3; 1.
 DR Pfam; PF01006; HCV NS4a; 1.
 DR Pfam; PF01001; HCV NS4b; 1.
 DR Pfam; PF01506; HCV NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RBP; 1.
 DR Pfam; PF01862; HCV NS1; 1.
 DR Pfam; PF01862; HCV NS1; 1.
 DR SMART; SMO0487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KM Hydrolyase; Nonstructural protein; Polypeptide; Transmembrane.
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326801 MW; 9FEE3D1B93B7AA4B CRC64;

Query Match 96.6%; Score 1715; DB 12; Length 3010;
 Best Local Similarity 95.6%; Pred. No. 1,2e-137;
 Matches 326; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCVHPELIFDITKLLAIFGPIMLTQAGITKVPYVRAQGLIRACMLV 60
 DB VRGRDAIILLTCVHPELIFDITKLLAIFGPIMLTQAGITKVPYVRAQGLIRACMLV 925
 QY 61 RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 120
 DB RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 985
 QY 926 RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 180
 DB RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 1045
 QY 121 WGAADTAACGDIISGLPVARSRRRIILGPADNFBGQGRLLAPITAYSQTRGLGCIIT 240
 DB WGAADTAACGDIISGLPVARSRRRIILGPADNFBGQGRLLAPITAYSQTRGLGCIIT 1105
 QY 181 SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 240
 DB SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 1105
 QY 1046 SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 1105
 DB SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 1105
 QY 241 CDLVGMQAPPGARSMTCTCGSSDLYLVTRHADYIPVRRGDSRGLSLSPRYSYLKSS 300
 DB CDLVGMQAPPGARSMTCTCGSSDLYLVTRHADYIPVRRGDSRGLSLSPRYSYLKSS 1165
 QY 1106 LDVGMQAPPGARSMTCTCGSSDLYLVTRHADYIPVRRGDSRGLSLSPRYSYLKSS 1165
 DB LDVGMQAPPGARSMTCTCGSSDLYLVTRHADYIPVRRGDSRGLSLSPRYSYLKSS 1165
 QY 301 GGPIICPSGHAIVGFRAAVCTRGAKAVDPIVESMETTR 341
 DB GGPIICPSGHAIVGFRAAVCTRGAKAVDPIVESMETTR 1206

RESULT 7
 Q9J3F4 PRELIMINARY; PRT; 3008 AA.

AC Q9J3F4
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polypeptide.
 GN MD34.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 RN NCBI_TaxID=11103;
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD34.
 RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RT "Characteristics of hepatitis C viral genome associated with disease
 progression.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN C AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; AF208024; AAF61205.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; PS0329; PS0329.
 DR HSP; P26663; IUXP.

DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005524; F: ATP binding; IEA.
 DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO; GO:0005487; F: electron transporter activity; IEA.
 DR GO; GO:0016787; F: hydrolase activity; IEA.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR GO; GO:0016740; F: transferase activity; IEA.
 DR GO; GO:0006118; F: electron transport; IEA.
 DR GO; GO:0006508; F: proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; F: transcription; IEA.
 DR GO; GO:0019079; F: viral genome replication; IEA.
 DR GO; GO:0019087; F: viral transformation; IEA.
 DR Interpro; IPR009003; Cys Ser trypsin.
 DR Interpro; IPR000345; CysC_heme_Bs.
 DR Interpro; IPR001410; DEAD.
 DR Interpro; IPR002522; HCV_capsid.
 DR Interpro; IPR002521; HCV_core.
 DR Interpro; IPR002519; HCV_env.
 DR Interpro; IPR002531; HCV_NS1.
 DR Interpro; IPR002518; HCV_NS2.
 DR Interpro; IPR000745; HCV_NS4a.
 DR Interpro; IPR001490; HCV_NS4b.
 DR Interpro; IPR002868; HCV_NS5a.
 DR Interpro; IPR002166; HCV_RBP.
 DR Interpro; IPR001650; Helicase_C.
 DR Interpro; IPR004109; Peptidase_C29.
 DR Interpro; IPR007094; RNA pol_DS_PS.
 DR Interpro; IPR007094; RNA pol_PSVlr.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR Pfam; PF00998; Viral_RBP; 1.
 DR Pfam; PF01862; HCV NS1; 1.
 DR Pfam; PF01862; HCV NS1; 1.
 DR SMART; SMO0487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KM Hydrolyase; Nonstructural protein; Polypeptide; Transmembrane.
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3008 AA; 326834 MW; 99AE09E14C3109F4 CRC64;

Query Match 96.5%; Score 1715; DB 12; Length 3008;
 Best Local Similarity 95.6%; Pred. No. 2.1e-137;
 Matches 326; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGRDAIILLTCVHPELIFDITKLLAIFGPIMLTQAGITKVPYVRAQGLIRACMLV 60
 DB VRGRDAIILLTCVHPELIFDITKLLAIFGPIMLTQAGITKVPYVRAQGLIRACMLV 923
 QY 61 RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 120
 DB RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 983
 QY 926 RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 180
 DB RKAAGHYVQMAFKLALGTYYVDHLPQDAHAGRLDAVAEPVFSMEVYKIT 1043
 QY 121 WGAADTAACGDIISGLPVARSRRRIILGPADNFBGQGRLLAPITAYSQTRGLGCIIT 240
 DB WGAADTAACGDIISGLPVARSRRRIILGPADNFBGQGRLLAPITAYSQTRGLGCIIT 1103
 QY 181 SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 240
 DB SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 1103
 QY 1044 SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 1103
 DB SLTRDKNQVEGEVQVSTATQSFATCNGVCMTHGASGKTLGPKPIQMYTNVD 1103
 QY 241 CDLVGMQAPPGARSMTCTCGSSDLYLVTRHADYIPVRRGDSRGLSLSPRYSYLKSS 300


```

Db      1104 ODLVGMQAPFGARSLFPCGSSDLYLTVRHADVIPRRRGDSRGLSPRVSYLKSS 1163
QY      301 GGPLLPSGSHAVGIFRAVCTRGVAKAVDFVVESEMTTMR 341
Db      1164 GGPLLPSGSHAVGIFRAVCTRGVAKAVDFVVESEMTTMR 1204

RESULT 8
ID      09JTH3      PRELIMINARY;      PRT; 3010 AA.
AC      09JTH3;
DT      01-OCT-2000 (TReMBLrel. 15, Created)
DT      01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MD19;
RA      Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT      "Characteristics of hepatitis C viral genome associated with disease
RT      progression.";
RL      Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND RNA. (BY SIMILARITY).
DR      EMBL; AF207760; AAF65950.1; -.
DR      PIR; A61196; A61196.
DR      HSP; P26663; IUXP.
DR      GO; GO:0016021; C:Integral to membrane; IEA.
DR      GO; GO:0019028; C:viral capsid; IEA.
DR      GO; GO:0019031; C:viral envelope; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR      GO; GO:0005489; F:electron transporter activity; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR      GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR      GO; GO:0006350; P:transcription; IEA.
DR      GO; GO:0019079; P:viral genome replication; IEA.
DR      GO; GO:0019087; P:viral transformation; IEA.
DR      InterPro; IPR009003; Cys_Ser_tryptsin.
DR      InterPro; IPR000345; CytC_heme_BS.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR002522; HCV_capsid.
DR      InterPro; IPR002521; HCV core.
DR      InterPro; IPR002519; HCV env.
DR      InterPro; IPR002531; HCV_NS1.
DR      InterPro; IPR002518; HCV_NS2.
DR      InterPro; IPR000745; HCV_NS4a.
DR      InterPro; IPR001490; HCV_NS4b.
DR      InterPro; IPR002868; HCV_NS5a.
DR      InterPro; IPR002166; HCV_RdRp.
DR      InterPro; IPR001650; Helicase_C.
DR      InterPro; IPR004109; Peptidase_C9.
DR      InterPro; IPR007095; RNA_pol_DS_PS.
DR      InterPro; IPR007094; RNA_pol_PSVLr.
DR      Pfam; PF01543; HCV_capsid; 1.
DR      Pfam; PF01542; HCV_core; 1.
DR      Pfam; PF01539; HCV_env; 1.
DR      Pfam; PF01560; HCV_NS1; 1.
DR      Pfam; PF01538; HCV_NS2; 1.
DR      Pfam; PF02907; HCV_NS3; 1.

```

```

DR      Pfam; PF01006; HCV_NS4a; 1.
DR      Pfam; PF01001; HCV_NS4b; 1.
DR      Pfam; PF01500; HCV_NS5a; 1.
DR      Pfam; PF00271; helicase_C; 1.
DR      Pfam; PF00998; viral_RdRp; 1.
DR      ProDom; PD186062; HCV_NS1; 1.
DR      SMART; SM00487; DEXDC; 1.
DR      PROSITE; PS00190; CYTOCHROME_C_1.
KW      Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW      Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ      SEQUENCE 3010 AA; 327234 MW; 44C34677649CB8DD CRC64;

Query Match      96.5%; Score 1715; DB 12; Length 3010;
Best Local Similarity 94.4%; Pred. No. 2,1e-137;
Matches 322; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

QY      1 VRGGRDAITLITCAVHPELIPDITKLLAIFGLMVLQAGITKVPFVRAQGLIRACMLV 60
Db      866 VAGGRDAITLITCAVHPELIPDITKLLAIFGLMVLQAGITRVPFVRAQGLIRACMLA 925
QY      61 RRAAGGHVQMAFMKLAITGTYVDHTLPLODAHAAGRLDAVAVEVITSDEVKIIT 120
Db      926 RRVAGGHVQMAFMKLAITGTYVDHTLPLODAHAAGRLDAVAVEVITSDMEKTIIT 985
QY      121 WGAADRAAGDITSGPVSARRGREITLGPADNFEQGRLLAPITRAYSQQTGILGCIIT 180
Db      986 WGAADRAAGDITLGLPVSARRKEIFLGPADNLEQGRLLAPITRAYSQQTGILGCIIT 1045
QY      181 SLTGRDKQVEGEVQVSTATQSPLATCVNGVCTVFHAGSKTLAGEKPIITQMTYND 240
Db      1046 SLTGRDQVGEVQVSTATQSPLATCVNGVCTVFHAGAKTLAGEKPIITQMTYND 1105
QY      241 ODLVGMQAPFGARSLFPCGSSDLYLTVRHADVIPVRRRGDSRGLSPRVSYLKSS 300
Db      1106 ODLVGMQSPGARSLFPCGSSDLYLTVRHADVIPVRRRGDSRGLSPRVSYLKSS 1165
QY      301 GGPLLPSGSHAVGIFRAVCTRGVAKAVDFVVESEMTTMR 341
Db      1166 GGPLLPSGSHAVGIFRAVCTRGVAKAVDFVVESEMTTMR 1206

RESULT 9
ID      09D7D7      PRELIMINARY;      PRT; 3010 AA.
AC      09D7D7;
DT      01-MAR-2001 (TReMBLrel. 16, Created)
DT      01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT      01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_TaxID=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=HCV1217;
RA      Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA      Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA      Mishiura S.;
RT      "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT      with hepatocellular carcinoma: the 'progression score' revisited.";
RL      Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC      -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND RNA. (BY SIMILARITY).
DR      EMBL; AB049100; BAB1813.1; -.
DR      PIR; A61196; A61196.
DR      PIR; PQ0804; PQ0804.
DR      PIR; PS0329; PS0329.
DR      HSP; P26663; IUXP.
DR      MEROPS; S29.002; -.
DR      GO; GO:0016021; C:Integral to membrane; IEA.

```

DR GO: 00019028; C: viral capsid; IEA.
 DR GO: 00019031; C: viral envelope; IEA.
 DR GO: 00005524; F: ATP binding; IEA.
 DR GO: 00008026; F: ATP dependent helicase activity; IEA.
 DR GO: 00005489; F: electron transporter activity; IEA.
 DR GO: 00003723; F: RNA binding; IEA.
 DR GO: 00003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: 00008236; F: serine-type peptidase activity; IEA.
 DR GO: 00005198; F: structural molecule activity; IEA.
 DR GO: 00016740; F: transferase activity; IEA.
 DR GO: 00006118; F: electron transport; IEA.
 DR GO: 00006508; F: proteolysis and peptidolysis; IEA.
 DR GO: 00006350; F: transcription; IEA.
 DR GO: 00019079; P: viral genome replication; IEA.
 DR GO: 00019087; P: viral transformation; IEA.
 DR InterPro: IPR003003; Cys Ser trypsin.
 DR InterPro: IPR003045; Cys_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR004109; peptidase C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF01506; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF00271; HCV NS5a; 1.
 DR Pfam: PF00998; viral RdRp; 1.
 DR Pfam: PD186062; HCV NS1; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR Coats protein; Envelope protein; Glycoprotein; Nonstructural protein;
 DR Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326909 MW; 5505C62EB0DA0519 CRC64;

Query Match 96.5%; Score 1715; DB 12; Length 3010;
 Best Local Similarity 95.3%; Pred. No. 2.1e-137;
 Matches 325; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

DR 1 VRGGRDAIILITCAVHPELFDITKLLAIFGFLMVLQAGITKVPFVBAQGLIRACMLV 60
 DB 866 VRGGRDAIILITCAVHPELFDITKLLAIFGFLMVLQAGITKVPFVBAQGLIRACMLV 925
 QY 61 RKAAGHTVQMAFMKLAALTGTYYVDHLPLOMANAGLDLAANVAVPVSFMEVXIIT 120
 DB 926 RKAAGHTVQMAFMKLAALTGTYYVDHLPLOMANAGLDLAANVAVPVSFMEVXIIT 985
 QY 121 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAITYTSQOTRGLLCIIT 180
 DB 986 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAITYTSQOTRGLLCIIT 1045
 QY 181 SLTGRDNQVEGEVQVSTATGSLATCVAGVCMVTFHAGSKTLAGPKPIITOMYTNV 240
 DB 1046 SLTGRDNQVEGEVQVSTATGSLATCVAGVCMVTFHAGSKTLAGPKPIITOMYTNV 1105
 QY 241 QDLVGNQAPRGASMTCTCGSSDLYVPHADIVVRRGRGSLSPRPVSYLKGSS 300
 DB 1106 QDLVGNQAPRGASMTCTCGSSDLYVPHADIVVRRGRGSLSPRPVSYLKGSS 1165

QY 301 GSPILCPGSHAVGIFRAVACTRGVAKAVDFIPVESMETTR 341
 DB 1166 GSPILCPGSHAVGIFRAVACTRGVAKAVDFIPVESMETTR 1206

RESULT 10
 ID 0807P3 PRELIMINARY: PRT, 3010 AA.
 AC 0807P3.
 DT 01-JUN-2003 (TRENDEL 24, Created)
 DT 01-JUN-2003 (TRENDEL 24, Last sequence update)
 DT 01-OCT-2003 (TRENDEL 25, Last annotation update)
 DE Polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI:taxid=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MILE;
 RX MEDLINE=22047193; PubMed=12051758;
 RA Kishine H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,
 RA Nio Y., Hosaka M., Miyazaki Y., Shimotohno K.,
 RT "Subgenomic replicon derived from a cell line infected with the
 RT hepatitis C virus.",
 RT Hepatitis C virus.",
 RT Biochem. Biophys. Res. Commun. 293:993-999(2002).
 RL EMBL, AB080299; BAC54896.1; -;
 DR GO: 00019028; C: viral capsid; IEA.
 DR GO: 00019031; C: viral envelope; IEA.
 DR GO: 00005524; F: ATP binding; IEA.
 DR GO: 00008026; F: ATP dependent helicase activity; IEA.
 DR GO: 00005489; F: electron transporter activity; IEA.
 DR GO: 00003723; F: RNA binding; IEA.
 DR GO: 00003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: 00008236; F: serine-type peptidase activity; IEA.
 DR GO: 00006118; F: structural molecule activity; IEA.
 DR GO: 00006508; F: proteolysis and peptidolysis; IEA.
 DR GO: 00006350; F: transcription; IEA.
 DR GO: 00019079; P: viral genome replication; IEA.
 DR GO: 00019087; P: viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser trypsin.
 DR InterPro: IPR003045; Cys_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR004109; peptidase C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; viral RdRp; 1.
 DR ProDom: PD186062; HCV NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00487; HELIC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR Polyprotein.

SQ SEQUENCE 3010 AA; 327097 MW; EE6418C7A723EE66 CRC64;
 Query Match 96.5%; Score 1715; DB 12; Length 3010;
 Best Local Similarity 95.6%; Pred. No. 2,1e-137;
 Matches 326; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGRDAILLTCAVHPLEIFDITKLLAIFGPMVLAQGITKYPVRAOGLIRACMLV 60
 DB 866 IRGRDAILLTCAVHPLEIFDITKLLAIFGPMVLAQGITKYPVRAOGLIRACMLV 925
 QY 61 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIT 120
 DB 926 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIT 985
 QY 121 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 180
 DB 986 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 1045
 QY 181 SLTGRDKNQVGEVQVSTATQSFATCVNGCWTVPFHGAGSKTLAPKGPITQMYTND 240
 DB 1046 SLTGRDKNQVGEVQVSTATQSFATCVNGCWTVPFHGAGSKTLAPKGPITQMYTND 1105
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300
 DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 1165
 QY 301 GGPLPCPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 341
 DB 1166 GGPLPCPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 11
 P88803 PRELIMINARY; PRT: 3010 AA.
 AC P88803
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV-1b;
 RX MEDLINE=95340824; PubMed=7542279;
 RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T., Yamamoto C., Izumi N., Maruno F., Sato C.;
 RT "Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b.";
 RL J. Clin. Invest. 96:224-230 (1995).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL: D50484; BAA09075.1; -.
 DR EMBL; A61196; A61196.
 DR HSP; P26663; INS3.
 DR GO: GO:0016021; C: integral to membrane; IEA.
 DR GO: GO:0019028; C: viral capsid; IEA.
 DR GO: GO:0019031; C: viral envelope; IEA.
 DR GO: GO:0005524; F: ATP binding; IEA.
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO: GO:0003723; F: RNA binding; IEA.
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F: structural molecule activity; IEA.

DR GO: GO:0016740; F: transferase activity; IEA.
 DR GO: GO:0006508; F: proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; F: transcription; IEA.
 DR GO: GO:0019079; F: viral genome replication; IEA.
 DR GO: GO:0019087; F: viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NSI.
 DR InterPro: IPR000745; HCV_NS4A.
 DR InterPro: IPR001450; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007035; RNA_pol_D5_PS.
 DR InterPro: IPR007034; RNA_pol_Pevir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NSI; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4A; 1.
 DR Pfam: PF01001; HCV_NS4B; 1.
 DR Pfam: PF01506; HCV_NS5A; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR SMART: SM00487; DEXDC1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; Transferase; Transmembrane.
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1972
 FT CHAIN 1973 2419
 FT CHAIN 2420 3010
 FT CHAIN NSSB.
 SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783FEF7B8 CRC64;
 Query Match 96.4%; Score 1714; DB 12; Length 3010;
 Best Local Similarity 94.7%; Pred. No. 2,6e-137;
 Matches 323; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 VRGRDAILLTCAVHPLEIFDITKLLAIFGPMVLAQGITKYPVRAOGLIRACMLV 60
 DB 866 VRGRDAILLTCAVHPLEIFDITKLLAIFGPMVLAQGITKYPVRAOGLIRACMLV 925
 QY 61 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIT 120
 DB 926 RKAAGHYVQMAFKMLAALTGYVVDHLTPLODMAHAGLRDLAFAVEPVIFSDMEVKIT 985
 QY 121 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 180
 DB 986 WGAADTAACGDIISGLPVSARRGREILLGPADNFEQGRLLAPITAYSOQTRGLGCIIT 1045
 QY 181 SLTGRDKNQVGEVQVSTATQSFATCVNGCWTVPFHGAGSKTLAPKGPITQMYTND 240
 DB 1046 SLTGRDKNQVGEVQVSTATQSFATCVNGCWTVPFHGAGSKTLAPKGPITQMYTND 1105
 QY 241 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 300
 DB 1106 QDLVGMQAPPGARSMPTCTCGSSDLVYVTRHADVIPIVRRGDSRGLSPRPVSYLKSS 1165
 QY 301 GGPLPCPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 341
 DB 1166 GGPLPCPSGHAHVGFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RA	Hijikata M.;
RL	Submitted (DEC-1993) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; D11397, BAA20975.1; .
DR	PIR; A61196; A61196.
DR	PIR; PS0329; PS0329.
DR	PDB; 1DXE; 28-MAR-02.
DR	GO; GO:0005524; F:ATP binding; IEA.
DR	GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR	GO; GO:0016787; F:hydrolyase activity; IEA.
DR	GO; GO:0003676; F:nucleic acid binding; IEA.
DR	GO; GO:0008286; F:serine-type peptidase activity; IEA.
DR	GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR	GO; GO:0019087; P:viral transformation; IEA.
DR	InterPro; IPR009003; Cys_Ser_cleypsin.
DR	InterPro; IPR003410; DEAD.
DR	InterPro; IPR002518; HCV_NS2.
DR	InterPro; IPR000745; HCV_NS4b.
DR	InterPro; IPR001490; HCV_NS4d.
DR	InterPro; IPR001650; Helicase_C.
DR	InterPro; IPR004109; Peptidase_C29.
DR	Pfam; PF01538; HCV_NS2; 1.
DR	Pfam; PF02907; HCV_NS3; 1.
DR	Pfam; PF01006; HCV_NS4; 1.
DR	Pfam; PF01001; HCV_NS4b; 1.
DR	Pfam; PF00271; helicase_C; 1.
DR	SMART; SW00487; DEXdc; 1.
DR	ATP-binding; Helicase; Hydrolase.
FT	KW NON TER
SQ	SEQUENCE 1186 AA; 126280 MW; 3417047BBA23729A CRC64;
Query Match	96.2%; Score 1711; DB 12; Length 1186;
Best Local Similarity	95.3%; Pred. No. 1,4e-137;
Matches 325; Conservative	9; Mismatches 7; Indels 0; Gaps 0;
OY	1 VRGRDRIILLTCAVHPELIFFDIKLLLAIFGPMLVTAQGITYXVPYPAQGLIRACMLV 60
Db	144 VRGRDAIILLTCVHPELIFFDIKLLLAIGPMIVTAQGMTRPVFPFAAGLRACMLV 203
OY	61 RKAAGHYVQMAEFKKALNLCTYYVDHLTPLODAAHGRLPLAVAVEFVPSDMDEVKIT 120
Db	204 RKVAGHYVQMAEFKKALNLCTYYVDHLTPLRDAHGLRDPLAVAEFVPSDMETKIT 263
OY	121 WGAOTTAACGDIIISGLPYSAARGREIILGPADNFEQGGRLAPITAYSQOTRGILGCIT 180
Db	264 WGADTTAACGDIISGLPYSAARGKEIILGPADSFGEGWRLLAPITAYSQOTRGILGCIT 323
OY	181 SLTRDRKNQVEGEVQVNSTAQSLATLCVNGVCMTVFHGASKTLAGKPGITOMYTND 240
Db	324 SLTRDRKNQVDEGEVQVNSTAQSLATLCVNVCMTVTHGASKTLAGKPGITOMYTND 383
OY	241 QDLVGMOAPGARSMPTCTCGSDLYLVTRHADVIPLYRRGDSRGSLSPRPVSYLKSS 300
Db	384 QDLVGMPAPPARKSMTPTCTCGSDLYLVTRHADVIPLYRRRGDSGSLSPRPVSYLKSS 443
OY	301 GGPLLCPBGHAVGIFRAAVCTRGYAKAVDPFVPSMETMTR 341
Db	444 GGPLLCPBGHVAGIFRAAVCTRGYAKAVDFIPVSEMETMTR 484
RESULT 13	
Q81817	PRELIMINARY; PRT; 2284 AA.
AC	Q81817;
DT	01-NOV-1996 (TREMBLrel_01, Created)
DT	01-NOV-1996 (TREMBLrel_01, Last sequence update)
DT	01-OCT-2003 (TREMBLrel_25, Last annotation update)
OS	Polyprotein precursor (Genome polypeptide).
DE	Hepatitis C virus type 2.
CC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC	Hepadnavirus.
OX	NCBI_TaxId=40271;
RN	[1]
RP	SEQUENCE FROM N.A.

RX MEDLINE=94068484; PubMed=7504283;
 RA Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,
 RA Akagi T., Kimura K., Shimotohno K.;
 RT "Proteolytic processing and membrane association of putative
 RT nonstructural proteins of hepatitis C virus.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).
 RL [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=94333810; PubMed=8056334;
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
 RT "Identification of the domain required for trans-cleavage activity of
 RT hepatitis C viral serine proteinase.";
 RL Gene 145:215-219(1994).
 RN [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=95056078; PubMed=7966638;
 RA Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
 RT "Hepatitis C virus polyprotein processing: kinetics and mutagenic
 RT analysis of serine proteinase-dependent cleavage.";
 RL J. Virol. 68:8418-8422(1994).
 RN [4]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=95185583; PubMed=7853491;
 RA Tanji Y., Hijikata M., Satoh S., Kaneko T., Shimotohno K.;
 RT "Hepatitis C virus-encoded nonstructural protein NS4A has versatile
 RT functions in viral protein processing.";
 RL J. Virol. 69:1575-1581(1995).
 DR EMBL, D16435; EAA03905.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P03329; P03329.
 DR HSSP; P26663; IUXP.
 DR GO; GO:0019012; C:viral protein processing; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003773; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral transformation; IEA.
 DR InterPro; IPR009003; Cys_Ser_lysin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4A.
 DR InterPro; IPR001490; HCV_NS4B.
 DR InterPro; IPR002868; HCV_NS5A.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01538; HCV_NS2_1.
 DR Pfam; PF02907; HCV_NS3_1.
 DR Pfam; PF01006; HCV_NS4a_1.
 DR Pfam; PF01001; HCV_NS4b_1.
 DR Pfam; PF01506; HCV_NS5a_1.
 DR Pfam; PF00271; helicase; 1.
 DR Pfam; PF00998; Viral_RdRp_1.
 DR SMART; SM00487; DEXDC; 1.
 KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
 FT SIGNAL; 4
 FT SIGNAL; 20
 FT SIGNAL; 55
 FT CHAIN; 21
 FT CHAIN; 84
 FT CHAIN; 301
 FT CHAIN; 931
 FT CHAIN; 985
 FT CHAIN; 986
 FT CHAIN; 1247
 FT CHAIN; 1693
 FT CHAIN; 2284

SQ SEQUENCE 2284 AA; 247213 MW; DC272A1517046337 CRC64;
 Query Match 96.2%; Score 1711; DB 12; Length 2284;
 Best Local Similarity 95.3%; Pred. No. 3,2e-137;
 Matches 325; Conservative 9; Mismatches 7; Indels 0; Gaps 0;
 QY 1 VRGRDAIILLTCVAHPELFDITKLLAI FGPMLYLQGITKVPYFAVAGLIRACMLV 60
 DB 140 VRGRDAIILLTCVAHPELFDITKLLAI FGPMLYLQGITKVPYFAVAGLIRACMLV 199
 QY 61 RKAGAGHYQMAFMKALATGYVDHLPDMDHAGRLDAVNEPTFEDMEKITT 120
 DB 200 RKAGAGHYQMAFMKALATGYVDHLPDMDHAGRLDAVNEPTFEDMEKITT 259
 QY 121 WGAADTAACDIIISGLPVSARRGRILLGPADNFEQGMRILAPITAYSOOTRGLGCIIT 180
 DB 260 WGAADTAACDIIISGLPVSARRGRILLGPADNFEQGMRILAPITAYSOOTRGLGCIIT 319
 QY 181 SLTRGRDKNVBEQVYVSTATQSPFLATCNGCMTVFHAGSKITLAGPGPITQMTYND 240
 DB 320 SLTRGRDKNVBEQVYVSTATQSPFLATCNGCMTVFHAGSKITLAGPGPITQMTYND 379
 QY 241 QDLVGMQAPPGARSMTPCTGSSDLYLVTRHADVPVRRGDSRGLSPRVSYLKSS 300
 DB 380 QDLVGMQAPPGARSMTPCTGSSDLYLVTRHADVPVRRGDSRGLSPRVSYLKSS 439
 QY 301 GEPFLCPGSHAVGIFRAAVCTRGVAKAVDFIPESMETTMR 341
 DB 440 GEPFLCPGSHAVGIFRAAVCTRGVAKAVDFIPESMETTMR 480
 RESULT 14
 ID Q68788 PRELIMINARY; PRT; 3010 AA.
 AC Q68788;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE HCV polyprotein (Genome polyprotein).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96362158; PubMed=8720135;
 RA Seki M., Honda Y.;
 RT "Phosphorocholate antisense oligodeoxynucleotides capable of
 RT inhibiting Hepatitis C virus gene expression: In vitro translation
 RT assay.";
 RL J. Biochem. 118:1199-1204(1995).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; D45172; BAA08120.1; -.
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P03329; P03329.
 DR HSSP; P26663; IUXP.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003773; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; P:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.

DR GO:0019087; P.viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_lypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002533; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVlr.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF0186062; HCV_NS1; 1.
 DR Pfam: PF00487; DEAD; 1.
 DR SMART: SM00487; DEAD; 1.
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; Transferrase; Transmembrane.
 DR Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.
 DR SEQUENCE 3010 AA; 326880 MW; EED840E6A505DE766 CRC64;

Query Match 96.2%; Score 1711; DB 12; Length 3010;
 Best local similarity 94.7%; Pred. No. 4.6e-137;
 Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

DR 1 VRGRDAIILITCAVHELIIFDITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLV 60
 DB 866 VRGRDAIILITCAVHELIIFDITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLV 925
 QY 61 RKAAGHYVOMAFKMAALGTCTYDHLTFLQDMANGLFDLAVANEPVFSMEYKITT 120
 DB 926 WKAAGHYVOMAFKMAALGTCTYDHLTFLQDMANGLFDLAVANEPVFSMEYKITT 965
 QY 121 WGAATTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
 DB 986 WGAATTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 1045
 QY 181 SLTGRDKNQVGEVQVAVSTATQSLATCGVGVCTVFGHAGSKTLAEPKPTITQMTTND 240
 DB 1046 SLTGRDKNQVGEVQVAVSTATQSLATCGVGVCTVFGHAGSKTLAEPKPTITQMTTND 1105
 QY 241 ODIVGMAAPGARSMTCCTGSSDLYLVTRHADVIPRRRGDSRGLSPRPVSYLKSS 300
 DB 1106 ODIVGMAAPGARSMTCCTGSSDLYLVTRHADVIPRRRGDSRGLSPRPVSYLKSS 1165
 QY 301 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETTKR 341
 DB 1166 GGPLCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETTKR 1206
 RESULT 15
 P89966 PRELIMINARY; PRT; 3010 AA.
 AC P89966;
 DT 01-MAY-1997 (Tremblrel. 03. Created)
 DT 01-MAY-1997 (Tremblrel. 03. Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25. Last annotation update)
 DR RNA for polypotein (Genome polypotein).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis C virus.

OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-type 1b;
 RA Tanaka T.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-type 1b;
 RA TANAKA T.;
 RT "TORP";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MENA (BY SIMILARITY).
 DR EMBL: D89872; BAA14035.1; ..
 DR PIR: A61196; A61196.
 DR PIR: P0246; P0246.
 DR PIR: P0804; P0804.
 DR PIR: P0329; P0329.
 DR HSP: P26653; IUXP.
 DR GO:0016021; C:Integral to membrane; IEA.
 DR GO:0019028; C:viral capsid; IEA.
 DR GO:0019031; C:viral envelope; IEA.
 DR GO:0005524; F:ATP binding; IEA.
 DR GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO:0003723; F:RNA binding; IEA.
 DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO:0005198; F:structural molecule activity; IEA.
 DR GO:0016740; F:transferase activity; IEA.
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO:0006350; P:transcription; IEA.
 DR GO:0019075; P:viral genome replication; IEA.
 DR GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_Ser_lypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002533; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVlr.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR Pfam: PF0186062; HCV_NS1; 1.
 DR Pfam: PF00487; DEAD; 1.
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein; RNA polymerase; Transferrase; Transmembrane.
 DR Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.
 DR SEQUENCE 3010 AA; 327023 MW; E075BD9CFDSD1261 CRC64;

Query Match 96.2%; Score 1711; DB 12; Length 3010;
 Best local similarity 95.3%; Pred. No. 4.6e-137;
 Matches 325; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

1 VRGRDAIILITCAVHELIIFDITKLLAIFGPIMLVLAQGITKVPYFVRAQGLIRACMLV 60

```

Db      866 VRGGRDAIILITCAVHPBELIFDIKLLALIGPLMVLAQWTRVFPVFAQGLIRACMLV 925
QY      61 RKAAGSHVQMAFMKLAALTSTVYDHLP:QDMAHAGLRDLAVAVEPIFSMEVKIIT 120
Db      926 RKNVGGHYVQMAFMKLAALTSTVYDHLP:LRDMAHAGLRDLAVAVEPIFSMETEKIIT 985
QY      121 WGAADTAACGDIISGLPVSARRGREILGPADNFEQGMRLIAPITAYSOOTRGILGCIIT 180
Db      986 WGAADTAACGDIISGLPVSARRGREILGPADNFEQGMRLIAPITAYSOOTRGILGCIIT 1045
QY      181 SLTGRDNQVGEVQVAVSTATQSSFLATCVNGVCWTFHAGSKTLAGPKGPIITQMTNV 240
Db      1046 SLTGRDNQVGEVQVAVSTATQSSFLATCVNGVCWTFHAGSKTLAGPKGPIITQMTNV 1105
QY      241 ODLVGMQAPPGARSMPTCTGSSDLVLTNRHADVIPVRRGDSRGSLLSPRPVSYLKSS 300
Db      1106 QDLVGMQAPPGARSMPTCTGSSDLVLTNRHADVIPVRRGDSRGSLLSPRPVSYLKSS 1165
QY      301 GGPILCPSGHAWGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db      1166 GGPILCPSGHAWGIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

```

Search completed: May 6, 2004, 09:35:49
 Job time : 32.8931 secs

Fri May 7 13:37:05 2004

us-10-650-585-14.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13.3837 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-14

Perfect score: 1778
Sequence: 1 VRGGRDAIILTCVHPELI.....RGVAKAVDFPVESEMTTMR 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625871 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCUTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1702	95.7	2201	4	US-09-539-601-6
2	1702	95.7	2201	4	US-09-539-601-15
3	1702	95.7	3010	4	US-09-539-601-3
4	1702	95.7	3010	4	US-09-539-601-21
5	1702	95.7	3010	4	US-09-539-601-27
6	1698	95.5	1692	3	US-09-263-933-4
7	1698	95.5	1692	3	US-09-263-933-2
8	1698	95.5	2307	3	US-09-263-933-1
9	1698	95.5	2307	3	US-09-263-933-2
10	1695	95.3	1692	3	US-09-263-933-11
11	1695	95.3	1692	3	US-09-263-933-9
12	1695	95.3	2307	3	US-09-263-933-9
13	1695	95.3	2307	4	US-09-919-901-9
14	1695	95.3	3010	4	US-09-539-601-13
15	1687	94.9	3010	3	US-09-014-416-3
16	1686	94.8	1692	3	US-09-263-933-18
17	1686	94.8	1692	3	US-09-263-933-18
18	1686	94.8	2307	3	US-09-263-933-16
19	1686	94.8	2307	3	US-09-263-933-16
20	1649	92.7	2013	1	US-08-324-977-12
21	1649	92.7	2013	2	US-08-384-616-12
22	1649	92.7	2013	2	US-08-904-686A-12
23	1649	92.7	2013	3	US-09-315-850-12
24	1649	92.7	2201	4	US-08-952-981A-2
25	1649	92.7	2620	1	US-08-324-977-32
26	1649	92.7	2620	2	US-08-384-616-32
27	1649	92.7	2620	2	US-08-904-686A-32

28	1649	92.7	2620	3	US-09-315-850-32	Sequence 32, Appl
29	1649	92.7	2621	1	US-08-324-977-36	Sequence 36, Appl
30	1649	92.7	2621	2	US-08-384-616-36	Sequence 36, Appl
31	1649	92.7	2621	2	US-08-904-686A-36	Sequence 36, Appl
32	1649	92.7	2621	3	US-09-315-850-36	Sequence 36, Appl
33	1649	92.7	3010	1	US-08-324-977-2	Sequence 2, Appl
34	1649	92.7	3010	1	US-08-324-977-14	Sequence 14, Appl
35	1649	92.7	3010	2	US-08-384-616-2	Sequence 2, Appl
36	1649	92.7	3010	2	US-08-384-616-14	Sequence 14, Appl
37	1649	92.7	3010	2	US-08-904-686A-2	Sequence 2, Appl
38	1649	92.7	3010	2	US-08-904-686A-14	Sequence 14, Appl
39	1649	92.7	3010	2	US-09-315-850-14	Sequence 14, Appl
40	1649	92.7	3010	3	US-09-315-850-14	Sequence 14, Appl
41	1570	88.3	1648	1	US-08-188-281B-12	Sequence 12, Appl
42	1570	88.3	1648	5	PCT-US94-07280-12	Sequence 12, Appl
43	1570	88.3	1648	5	PCT-US95-01087-12	Sequence 12, Appl
44	1570	88.3	3011	1	US-08-188-281B-1	Sequence 1, Appl
45	1570	88.3	3011	1	US-08-453-552-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-539-601-6
; Sequence 6, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartschlagel, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-6

Query Match 95.7%; Score 1702; DB 4; Length 2201;
Best Local Similarity 94.1%; Pred. No. 48-164;
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY	1	VRGGRPAIILTCVHPELI	FDITKLLAI	FGPLMVLQGITKVPYFVRAQGLIRACMLV	60	
Db	57	VRGGRPAVILTCVHPELI	FTITKLLAI	GLPLMVLQGITKVPYFVRAHGLIRACMLV	116	
QY	61	RKAAGHYQMAFMKLAAL	TGYVVDHL	FLQDMAHAGRLAVAVEPVSFDMEXKIT	120	
Db	117	RKAAGHYQMAFMKLAAL	TGYVVDHL	FLQDMAHAGRLDIAVAEPVFSMETKVIIT	176	
QY	121	WGADTAACDIIISGLPV	ARRRGREILL	SPADNFEQGMWLLAPITAYSCOTRGLGCIIT	180	
Db	177	WGADTAACDIIISGLPV	ARRRGREILL	SPADNFEQGMWLLAPITAYSCOTRGLGCIIT	236	
QY	181	SILGRDKNOVEBHVQV	STATOSFLAT	CNGCWMVYFHAGSKTLA	PGKPIITQWYTNVD	240
Db	237	SILGRDRNOVEBHVQV	STATOSFLAT	CNGCWMVYFHAGSKTLA	PGKPIITQWYTNVD	296
QY	241	QDIVGQAQPGARSMTP	CTCGSSDLYL	VTYRAHADVIVPRRGRGDSRGLSPREVSYLKSS	300	
Db	297	QDIVGQAQPGARSLTP	CTCGSSDLYL	VTYRAHADVIVPRRGRGDSRGLSPREVSYLKSS	356	
QY	301	GGPLCPSGHNAVIGIR	PAVCTRGVAKAVD	FVSESEMTTMR	341	
Db	357	GGPLCPSGHNAVIGIR	PAVCTRGVAKAVD	FVSESEMTTMR	397	

RESULT 2

RESULT 2

RESULT 5
US-09-539-601-27
; Sequence 27, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Barteneschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 27
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-27

Query Match 95.7%; Score 1702; DB 4; Length 3010;
Best Local Similarity 94.1%; Pred. No. 6.3e-164;
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 60
DB VRGGRDAVILITCAIHPELIFTITKLLALIGPLMLVQAGITVFPVFAOGLIRACMLV 925
QY 61 KKAAGHYVQMAFMKLAALGTGYVDHLTPLOMAHAGRLDAVAEPIVFSDEMKIIT 120
DB VRGGRDAVILITCAIHPELIFTITKLLALIGPLMLVQAGITVFPVFAOGLIRACMLV 925
QY 926 RKVAGHYVQMAFMKLAALGTGYVDHLTPLOMAHAGRLDAVAEPIVFSDEMKIIT 985
DB WADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 180
QY 121 WADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 180
DB WADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 1045
QY 181 SLGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 240
DB SLGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 1105
QY 1046 SLGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 1105
DB SLGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 1105
QY 241 QDLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 300
DB QDLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 1165
QY 1106 QDLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 1165
DB QDLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 1165
QY 301 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTR 341
DB GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

RESULT 6
US-09-263-933-4
; Sequence 4, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-4

Query Match 95.5%; Score 1698; DB 3; Length 1692;

Best Local Similarity 94.1%; Pred. No. 6.9e-164;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 61
DB RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 205
QY 146 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 205
DB RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 205
QY 62 KKAAGHYVQMAFMKLAALGTGYVDHLTPLOMAHAGRLDAVAEPIVFSDEMKIIT 121
DB KKAAGHYVQMAFMKLAALGTGYVDHLTPLOMAHAGRLDAVAEPIVFSDEMKIIT 265
QY 122 GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 181
DB GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 325
QY 266 GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 325
DB GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 385
QY 182 LTGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 241
DB LTGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 385
QY 326 LTGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 385
DB LTGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 385
QY 242 DLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 301
DB DLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 445
QY 386 DLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 445
DB DLVQWAPPGASMTPTCTGSSDLYLVTTHADVIPIRRRGDSRGLSPRPVSYLKSSG 445
QY 302 GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTR 341
DB GGPLCPSGHAAGIFRAAVCTRGVAKAVDFIPVESMETTR 465

RESULT 7
US-09-919-901-4
; Sequence 4, Application US/09919901
; Patent No. 6539738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: :
US-09-919-901-4

Query Match 95.5%; Score 1698; DB 4; Length 1692;
Best Local Similarity 94.1%; Pred. No. 6.9e-164;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 61
DB RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 205
QY 146 RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 205
DB RGRDAIILITCAVHPELIFDITKLLAIFGPIMLVQAGITVFPVFAOGLIRACMLV 205
QY 62 KKAAGHYVQMAFMKLAALGTGYVDHLTPLOMAHAGRLDAVAEPIVFSDEMKIIT 121
DB KKAAGHYVQMAFMKLAALGTGYVDHLTPLOMAHAGRLDAVAEPIVFSDEMKIIT 265
QY 122 GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 181
DB GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 325
QY 266 GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 325
DB GADTAACGDIILGLPVASARGREIILGPADSLGSGWRLLAPITAYSOOTRGLGCIIT 325
QY 182 LTGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 241
DB LTGRDNQYGEVQVSTATQSFATCVNGVCMVTFHAGSKTLGPKPITOMYTNDV 241

Db 326 LTGRDKNOVEGEVQVSTATOSFLATCVCNVCWTVYHAGSKTLAAGPKBITOMTNNVDQ 385
Qy 242 DLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIYVRRRGRDSSGSLSPRPVSYLKSSG 301
Db 386 DLVGMQAPPGARSLTPTCGSSDLVYVTRHADVIYVRRRGRDSSGSLSPRPVSYLKSSG 445
Qy 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 341
Db 446 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 485

RESULT 8

US-09-263-933-2

; Sequence 2, Application US/09263933

; Patent No. 6280940

; GENERAL INFORMATION:

; APPLICANT: Potts, Karen E.

; APPLICANT: Jackson, Roberta L.

; APPLICANT: Patrick, Amy K.

; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

; FILE REFERENCE: 0125-0005A

; CURRENT APPLICATION NUMBER: US/09/263,933

; EARLIER FILING DATE: 1999-03-08

; EARLIER FILING DATE: 1998-08-05

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2307

; TYPE: PRT

; ORGANISM: Artificial Sequence

US-09-263-933-2

Query Match 95.5%; Score 1698; DB 3; Length 2307;
Best Local Similarity 94.1%; Pred. No. 1.1e-163;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGGDAITLLTCAVHPFLIDITKLLAIFGPIMLVLAQGITKYFVFAQGLIRACMLVR 61
Db 238 RGGDAITLLTCAVHPFLIDITKLLAIFGPIMLVLAQGITKYFVFAQGLIRACMLVR 297
Qy 62 KAAGSHYQVAFMKLAALGTYYVDHLTPLODMNAHGLRDLAVAVEPVFSDEMKIITW 121
Db 298 KAAGSHYQVAFMKLAALGTYYVDHLTPLODMNAHGLRDLAVAVEPVFSDEMKIITW 357
Qy 122 GADTAAGCDITLIGPVSARRGRELILGPADNFEQGRLLAPITAYSQOIRGLGCIITS 181
Db 358 GADTAAGCDITLIGPVSARRGRELILGPADNFEQGRLLAPITAYSQOIRGLGCIITS 417
Qy 182 LTGRDKNOVEGEVQVSTATOSFLATCVCNVCWTVYHAGSKTLAAGPKBITOMTNNVDQ 241
Db 418 LTGRDKNOVEGEVQVSTATOSFLATCVCNVCWTVYHAGSKTLAAGPKBITOMTNNVDQ 477
Qy 242 DLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIYVRRRGRDSSGSLSPRPVSYLKSSG 301
Db 478 DLVGMQAPPGARSLTPTCGSSDLVYVTRHADVIYVRRRGRDSSGSLSPRPVSYLKSSG 537
Qy 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 341
Db 538 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 577

RESULT 9

US-09-919-901-2

; Sequence 2, Application US/09919901

; Patent No. 6599738

; GENERAL INFORMATION:

; APPLICANT: Potts, Karen E.

; APPLICANT: Jackson, Roberta L.

; APPLICANT: Patrick, Amy K.

; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: :
US-09-919-901-2

Query Match 95.5%; Score 1698; DB 4; Length 2307;
Best Local Similarity 94.1%; Pred. No. 1.1e-163;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGGDAITLLTCAVHPFLIDITKLLAIFGPIMLVLAQGITKYFVFAQGLIRACMLVR 61
Db 238 RGGDAITLLTCAVHPFLIDITKLLAIFGPIMLVLAQGITKYFVFAQGLIRACMLVR 297
Qy 62 KAAGSHYQVAFMKLAALGTYYVDHLTPLODMNAHGLRDLAVAVEPVFSDEMKIITW 121
Db 298 KAAGSHYQVAFMKLAALGTYYVDHLTPLODMNAHGLRDLAVAVEPVFSDEMKIITW 357
Qy 122 GADTAAGCDITLIGPVSARRGRELILGPADNFEQGRLLAPITAYSQOIRGLGCIITS 181
Db 358 GADTAAGCDITLIGPVSARRGRELILGPADNFEQGRLLAPITAYSQOIRGLGCIITS 417
Qy 182 LTGRDKNOVEGEVQVSTATOSFLATCVCNVCWTVYHAGSKTLAAGPKBITOMTNNVDQ 241
Db 418 LTGRDKNOVEGEVQVSTATOSFLATCVCNVCWTVYHAGSKTLAAGPKBITOMTNNVDQ 477
Qy 242 DLVGMQAPPGARSMTPCTCGSSDLVYVTRHADVIYVRRRGRDSSGSLSPRPVSYLKSSG 301
Db 478 DLVGMQAPPGARSLTPTCGSSDLVYVTRHADVIYVRRRGRDSSGSLSPRPVSYLKSSG 537
Qy 302 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 341
Db 538 GPLLCPGSHAVGIFRAAVCTRGVAKAVDFIVPESMETMTR 577

RESULT 10

US-09-263-933-11

; Sequence 11, Application US/09263933

; Patent No. 6280940

; GENERAL INFORMATION:

; APPLICANT: Potts, Karen E.

; APPLICANT: Jackson, Roberta L.

; APPLICANT: Patrick, Amy K.

; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

; FILE REFERENCE: 0125-0005A

; CURRENT APPLICATION NUMBER: US/09/263,933

; EARLIER FILING DATE: 1999-03-08

; EARLIER FILING DATE: 1998-08-05

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11

; LENGTH: 1692

; ORGANISM: Artificial Sequence

US-09-263-933-11

Query Match 95.3%; Score 1695; DB 3; Length 1692;
Best Local Similarity 93.8%; Pred. No. 1.4e-153;
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

```

QY 2 RGRDAIILITCAVHPELIDITKLLAIFGRLMTQAGITKVPYFRAAGILRACMLVR 61
DB 146 RGRDAIILITCAVHPELIDITKLLAIFGRLMTQAGITKVPYFRAAGILRACMLVR 205
QY 62 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITM 121
DB 206 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITM 265
QY 122 GADTAACGDIISGLPVASARRGRIILGPADNFBGQWRLAPITAYSOQTRGLGCIITS 181
DB 266 GADTAACGDIISGLPVASARRGRIILGPADNFBGQWRLAPITAYSOQTRGLGCIITS 325
QY 182 LTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 241
DB 326 LTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 385
QY 242 DLVGMQAPPGARSMTPTCTCGSSDLVLTTRHADVIPIVRRGDSRGSILSPRPVSYLKSSG 301
DB 386 DLVGMQAPPGARSMTPTCTCGSSDLVLTTRHADVIPIVRRGDSRGSILSPRPVSYLKSSG 445
QY 302 GPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341
DB 446 GPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 485

```

RESULT 11

```

US-09-919-901-11
; Sequence 11, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
; US-09-919-901-11

```

```

Query Match 95.3%; Score 1695; DB 4; Length 1692;
Best Local Similarity 93.8%; Pred. No. 1.4e-163;
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;
QY 2 RGRDAIILITCAVHPELIDITKLLAIFGRLMTQAGITKVPYFRAAGILRACMLVR 61
DB 146 RGRDAIILITCAVHPELIDITKLLAIFGRLMTQAGITKVPYFRAAGILRACMLVR 205
QY 62 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITM 121
DB 206 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITM 265
QY 122 GADTAACGDIISGLPVASARRGRIILGPADNFBGQWRLAPITAYSOQTRGLGCIITS 181
DB 266 GADTAACGDIISGLPVASARRGRIILGPADNFBGQWRLAPITAYSOQTRGLGCIITS 325
QY 182 LTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 241
DB 326 LTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 385
QY 242 DLVGMQAPPGARSMTPTCTCGSSDLVLTTRHADVIPIVRRGDSRGSILSPRPVSYLKSSG 301

```

```

DB 386 DLVGMQAPPGARSMTPTCTCGSSDLVLTTRHADVIPIVRRGDSRGSILSPRPVSYLKSSG 445
QY 302 GPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341
DB 446 GPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 485

```

RESULT 12

```

US-09-263-933-9
; Sequence 9, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; CURRENT FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRT
; ORGANISM: Artificial Sequence
; US-09-263-933-9

```

```

Query Match 95.3%; Score 1695; DB 3; Length 2307;
Best Local Similarity 93.8%; Pred. No. 2.2e-163;
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 RGRDAIILITCAVHPELIDITKLLAIFGRLMTQAGITKVPYFRAAGILRACMLVR 61
DB 238 RGRDAIILITCAVHPELIDITKLLAIFGRLMTQAGITKVPYFRAAGILRACMLVR 297
QY 62 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITM 121
DB 298 KAAGHYVQMAFMKLAALGTGYVDHLTPLODMAHAGRLDAVAEVPVFSMEVKIITM 357
QY 122 GADTAACGDIISGLPVASARRGRIILGPADNFBGQWRLAPITAYSOQTRGLGCIITS 181
DB 358 GADTAACGDIISGLPVASARRGRIILGPADNFBGQWRLAPITAYSOQTRGLGCIITS 417
QY 182 LTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 241
DB 418 LTGRDKNOVEGEVOVSTATOSFLATCVNGVCMVTFHAGSKTLAGPKGPIITOMYTNVDQ 477
QY 242 DLVGMQAPPGARSMTPTCTCGSSDLVLTTRHADVIPIVRRGDSRGSILSPRPVSYLKSSG 301
DB 478 DLVGMQAPPGARSMTPTCTCGSSDLVLTTRHADVIPIVRRGDSRGSILSPRPVSYLKSSG 537
QY 302 GPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341
DB 538 GPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 577

```

RESULT 13

```

US-09-919-901-9
; Sequence 9, Application US/09919901
; Patent No. 6599738
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/919,901
; CURRENT FILING DATE: 2001-08-02

```

```

; PRIOR APPLICATION NUMBER: 09/263,933
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 2307
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION:
US-09-919-901-9

```

```

Query Match      95.3%; Score 1695; DB 4; Length 2307;
Best Local Similarity 93.8%; Pred. No. 2.2e-163;
Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 RGGSDAIIILTCVHPELIFDITKLLAIFGPIMLVQAGITKPYFVRAQGLIRACMLV 61
DB 238 RGGSDAIIILTCVHPELIFDITKLLAIFGPIMLVQAGITKPYFVRAQGLIRACMLV 297
QY 62 KAAAGHYVQMAFMKLAALGTYYVDLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITW 121
DB 298 KAAAGHYVQMAFMKLAALGTYYVDLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIITW 357
QY 122 GADTAAAGDIIISGLPVARSRRREILGPADNFEQGGKRLAPITAYSOOTRGLIGCIIT 181
DB 358 GADTAAAGDIIISGLPVARSRRREILGPADNFEQGGKRLAPITAYSOOTRGLIGCIIT 417
QY 182 LTGSDKNQVGEVQVSTATOSFLATCNGVCWTVFHGAGSKTLAPKPGITOMYTNVDQ 241
DB 418 LTGSDKNQVGEVQVSTATOSFLATCNGVCWTVFHGAGSKTLAPKPGITOMYTNVDQ 477
QY 242 DLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRDSRGSILSPRPVSYLKGSSG 301
DB 478 DLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRDSRGSILSPRPVSYLKGSSG 537
QY 302 GPLPCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
DB 538 GPLPCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 577

```

```

RESULT 14
US-09-539-601-33
; Sequence 33; Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Barenscnager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 3010
; TYPE: PRF
; ORGANISM: Hepatitis C virus
US-09-539-601-33

```

```

Query Match      95.3%; Score 1695; DB 4; Length 3010;
Best Local Similarity 93.8%; Pred. No. 3.3e-163;
Matches 320; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 1 VRGRDAIIILTCVHPELIFDITKLLAIFGPIMLVQAGITKPYFVRAQGLIRACMLV 60
DB 866 VRGRDAIIILTCVHPELIFDITKLLAIFGPIMLVQAGITKPYFVRAQGLIRACMLV 925
QY 61 RKAAGHYVQMAFMKLAALGTYYVDLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIIT 120
DB 1166 RKAAGHYVQMAFMKLAALGTYYVDLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIIT 120

```

```

DB 926 RKAAGHYVQMAFMKLAALGTYYVDLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIIT 985
QY 121 WCADTAAAGDIIISGLPVARSRRREILGPADNFEQGGKRLAPITAYSOOTRGLIGCIIT 180
DB 986 WCADTAAAGDIIISGLPVARSRRREILGPADNFEQGGKRLAPITAYSOOTRGLIGCIIT 1045
QY 181 SLTGRDNQVGEVQVSTATOSFLATCNGVCWTVFHGAGSKTLAPKPGITOMYTNVD 240
DB 1046 SLTGRDNQVGEVQVSTATOSFLATCNGVCWTVFHGAGSKTLAPKPGITOMYTNVD 1105
QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRDSRGSILSPRPVSYLKGSS 300
DB 1106 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRDSRGSILSPRPVSYLKGSS 1165
QY 301 GPLPCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
DB 1166 GPLPCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

```

```

RESULT 15
US-09-014-416-3
; Sequence 3; Application US/09014416
; Patent No. 6153421
; GENERAL INFORMATION:
; APPLICANT: Yanagi, Masayuki
; APPLICANT: Buht, Jens
; APPLICANT: Emerson, Susanne U.
; APPLICANT: Purcell, Robert H.
; TITLE OF INVENTION: CLONED GENOMES OF INFECTIOUS HEPATITIS C VIRUSES AND
; FILE REFERENCE: 20264276
; CURRENT APPLICATION NUMBER: US/09/014,416
; CURRENT FILING DATE: 1998-01-27
; EARLIER APPLICATION NUMBER: US 60/053,062
; EARLIER FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3010
; TYPE: PRF
; ORGANISM: Hepatitis C virus
US-09-014-416-3

```

```

Query Match      94.9%; Score 1687; DB 3; Length 3010;
Best Local Similarity 93.3%; Pred. No. 2.2e-162;
Matches 318; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

```

```

QY 1 VRGRDAIIILTCVHPELIFDITKLLAIFGPIMLVQAGITKPYFVRAQGLIRACMLV 60
DB 866 VRGRDAIIILTCVHPELIFDITKLLAIFGPIMLVQAGITKPYFVRAQGLIRACMLV 925
QY 61 RKAAGHYVQMAFMKLAALGTYYVDLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIIT 120
DB 926 RKAAGHYVQMAFMKLAALGTYYVDLTPLODMAHAGLRDLAVAVEPVIFSDMEVKIIT 985
QY 121 WCADTAAAGDIIISGLPVARSRRREILGPADNFEQGGKRLAPITAYSOOTRGLIGCIIT 180
DB 986 WCADTAAAGDIIISGLPVARSRRREILGPADNFEQGGKRLAPITAYSOOTRGLIGCIIT 1045
QY 181 SLTGRDNQVGEVQVSTATOSFLATCNGVCWTVFHGAGSKTLAPKPGITOMYTNVD 240
DB 1046 SLTGRDNQVGEVQVSTATOSFLATCNGVCWTVFHGAGSKTLAPKPGITOMYTNVD 1105
QY 241 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRDSRGSILSPRPVSYLKGSS 300
DB 1106 QDLVGMQAPPGARSMTPTCGSSDLVYVTRHADVIPIRRRDSRGSILSPRPVSYLKGSS 1165
QY 301 GPLPCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
DB 1166 GPLPCPSGHAIVGIFRAAVCTRGVAKAVDFIPVESMETTR 1206

```

Search completed: May 6, 2004, 09:39:05

Fri May 7 13:37:05 2004

Job time : 14.3837 secs

us-10-650-585-14.ra1

Page 7

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 34.8831 Seconds
(without alignments)
2713.357 Million cell updates/sec

Title: US-10-650-585-14

Perfect score: 1778
Sequence: 1 VRGGRDAIILLTCVAHPELI.....RGVAKAVDFIPVSMETTR 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/PCRUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1778	100.0	341	13	US-10-017-736-14
2	1778	100.0	341	16	US-10-650-585-14
3	1778	100.0	352	13	US-10-017-736-13
4	1778	100.0	352	16	US-10-650-585-13
5	1778	100.0	380	13	US-10-017-736-12
6	1778	100.0	380	16	US-10-650-585-12
7	1778	100.0	393	13	US-10-017-736-11
8	1778	100.0	393	16	US-10-650-585-11
9	1778	100.0	403	13	US-10-017-736-2
10	1778	100.0	409	16	US-10-650-585-2
11	1702	95.7	2201	13	US-10-029-907-3
12	1702	95.7	2201	14	US-10-309-561-3
13	1702	95.7	3010	12	US-10-467-000-1
14	1698	95.5	1692	10	US-09-919-901-4
15	1698	95.5	1692	14	US-10-191-966-4

16	1698	95.5	2307	10	US-09-919-901-2	Sequence 2, Appli
17	1698	95.5	2307	14	US-10-191-966-2	Sequence 2, Appli
18	1695	95.3	1692	14	US-09-919-901-11	Sequence 11, Appli
19	1695	95.3	1692	14	US-10-191-966-11	Sequence 11, Appli
20	1695	95.3	2307	14	US-09-919-901-9	Sequence 9, Appli
21	1695	95.3	2307	14	US-10-191-966-9	Sequence 9, Appli
22	1686	94.8	1592	10	US-09-919-901-18	Sequence 18, Appli
23	1686	94.8	1592	14	US-10-191-966-18	Sequence 18, Appli
24	1686	94.8	2307	14	US-09-919-901-16	Sequence 16, Appli
25	1686	94.8	2307	10	US-10-191-966-16	Sequence 16, Appli
26	1649	92.7	2201	13	US-10-085-476-2	Sequence 2, Appli
27	1589	89.4	303	13	US-10-017-736-10	Sequence 10, Appli
28	1589	89.4	303	16	US-10-650-585-10	Sequence 10, Appli
29	1589	89.4	334	13	US-10-017-736-4	Sequence 4, Appli
30	1589	89.4	334	16	US-10-650-585-4	Sequence 4, Appli
31	1580	88.9	303	13	US-10-017-736-18	Sequence 18, Appli
32	1580	88.9	303	16	US-10-650-585-18	Sequence 18, Appli
33	1579	88.8	303	13	US-10-017-736-16	Sequence 16, Appli
34	1579	88.8	303	16	US-10-650-585-16	Sequence 16, Appli
35	1570	88.3	301	13	US-10-017-736-17	Sequence 17, Appli
36	1570	88.3	301	16	US-10-650-585-17	Sequence 17, Appli
37	1569	88.2	3011	9	US-09-742-659-4	Sequence 4, Appli
38	1569	88.2	3011	10	US-09-891-894-3	Sequence 3, Appli
39	1569	88.2	3011	14	US-10-184-150-3	Sequence 3, Appli
40	1569	88.2	3011	15	US-10-328-997-3	Sequence 3, Appli
41	1569	88.2	3012	9	US-09-238-076-2	Sequence 2, Appli
42	1569	88.2	3012	10	US-09-295-937-2	Sequence 2, Appli
43	1569	88.1	3012	10	US-09-917-563-2	Sequence 2, Appli
44	1566	88.1	3011	12	US-10-296-734-406	Sequence 406, App
45	1565	88.0	3011	9	US-09-916-359-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-017-736-14 Application US/10017736
; Sequence 14, Pufified Active HCV NS2/3 Protease
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Pufified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-14

Query Match	100.0%	Score 1778	DB 13	Length 341
Best Local Similarity	100.0%	Pred. No. 2.7e-174	Indels 0	Gaps 0
Matches 341	Conservative 0	Mismatches 0		
QY	1	VRGGRDAIILLTCVAHPELI	FDITKLLAIFGRLMWLGITKVFVFAOGIRACMLV	60
DB	1	VRGGRDAIILLTCVAHPELI	FDITKLLAIFGRLMWLGITKVFVFAOGIRACMLV	60
QY	61	RKAAGHYVQMAFMKLAALGTGVYDHLTFLDQWAGLRDLAAVEPVI	FSDEMYKIIT	120
DB	61	RKAAGHYVQMAFMKLAALGTGVYDHLTFLDQWAGLRDLAAVEPVI	FSDEMYKIIT	120
QY	121	WGADTLACGDIISGLFVSARREILLGPADNREGQMLLPITVSOOTGLGCIIT		180
DB	121	WGADTLACGDIISGLFVSARREILLGPADNREGQMLLPITVSOOTGLGCIIT		180
QY	181	SLRGDRKNQVEGEVQVSTATOSFLATCVNGVCMVIFHAGSKTLAGPKGITOMTVND		240

Db 181 SLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVTHGAGSKTLAQRKPIITQMTYND 240
 QY 241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTTHADVIPIVRRRDSRGSILSPRVSYLKSS 300
 Db 241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTTHADVIPIVRRRDSRGSILSPRVSYLKSS 300
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
 Db 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341

RESULT 2

US-10-650-585-14
 ; Sequence 14, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 341
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-650-585-14

Query Match 100.0%; Score 1778; DB 16; Length 341;
 Best Local Similarity 100.0%; Pred. No. 2.7e-174;
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 60
 Db 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 60
 QY 61 RKAAGHYVQMAFPMKLAALGTYYDHLTPLODMAHAGLBDLAAVEPVIFSDMEVKIIT 120
 Db 61 RKAAGHYVQMAFPMKLAALGTYYDHLTPLODMAHAGLBDLAAVEPVIFSDMEVKIIT 120
 QY 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180
 Db 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180
 QY 181 SLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVTHGAGSKTLAQRKPIITQMTYND 240
 Db 181 SLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVTHGAGSKTLAQRKPIITQMTYND 240
 QY 241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTTHADVIPIVRRRDSRGSILSPRVSYLKSS 300
 Db 241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTTHADVIPIVRRRDSRGSILSPRVSYLKSS 300
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
 Db 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341

RESULT 3

US-10-017-736-13
 ; Sequence 13, Application US/10017736
 ; Publication No. US20020192640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/017,736
 ; CURRENT FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 352
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-017-736-13

Query Match 100.0%; Score 1778; DB 13; Length 352;
 Best Local Similarity 100.0%; Pred. No. 2.9e-174;
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 60
 Db 12 VRGGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 71
 QY 61 RKAAGHYVQMAFPMKLAALGTYYDHLTPLODMAHAGLBDLAAVEPVIFSDMEVKIIT 120
 Db 72 RKAAGHYVQMAFPMKLAALGTYYDHLTPLODMAHAGLBDLAAVEPVIFSDMEVKIIT 131
 QY 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180
 Db 132 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 191
 QY 181 SLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVTHGAGSKTLAQRKPIITQMTYND 240
 Db 192 SLTGRDKNQVEGEVQVSTATQSFATCNGVCMTVTHGAGSKTLAQRKPIITQMTYND 251
 QY 241 QDLVGMQAPPGARSMTPTCTGSSDLYLVTTHADVIPIVRRRDSRGSILSPRVSYLKSS 300
 Db 252 QDLVGMQAPPGARSMTPTCTGSSDLYLVTTHADVIPIVRRRDSRGSILSPRVSYLKSS 311
 QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
 Db 312 GGPLCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 352

RESULT 4

US-10-650-585-13
 ; Sequence 13, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 352
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-650-585-13

Query Match 100.0%; Score 1778; DB 16; Length 352;
 Best Local Similarity 100.0%; Pred. No. 2.9e-174;
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 60
 Db 12 VRGGRDAIILITCAVHPELIFDITKLLAIFGRLMVLQAGITKVPYFVRAQGLIRACMLV 71
 QY 61 RKAAGHYVQMAFPMKLAALGTYYDHLTPLODMAHAGLBDLAAVEPVIFSDMEVKIIT 120
 Db 72 RKAAGHYVQMAFPMKLAALGTYYDHLTPLODMAHAGLBDLAAVEPVIFSDMEVKIIT 131
 QY 121 WGAADTAACGDIISGLPVARSRRREILGPADNFEQGMRLAPITAYSQOTRGLGCIIT 180

Fri May 7 13:37:05 2004

us-10-650-585-14.rapb

Page 3

```
Db 132 WGADTAACGDIISGLPVASARGREILLGPADNFEQGGWRLAPITAYSQOTRGLGCIIT 191
Qy 181 SLTGRDNQVGEVGVVSTATQSLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYND 240
Db 192 SLTGRDNQVGEVGVVSTATQSLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYND 251
Qy 241 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLSPPRVSYLKSS 300
Db 252 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLSPPRVSYLKSS 311
Qy 301 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db 312 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 352
```

```
RESULT 5
US-10-017-736-12
; Sequence 12, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-12
```

```
Query Match 100.0%; Score 1778; DB 13; Length 380;
Best Local Similarity 100.0%; Pred. No. 3,2e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VRGSDAIIILTCVAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACKLV 60
Db 40 VRGSDAIIILTCVAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACKLV 99
Qy 61 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMNHAQLRLAVAVEPIVFSMEVKIIT 120
Db 100 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMNHAQLRLAVAVEPIVFSMEVKIIT 159
Qy 121 WGADTAACGDIISGLPVASARGREILLGPADNFEQGGWRLAPITAYSQOTRGLGCIIT 180
Db 160 WGADTAACGDIISGLPVASARGREILLGPADNFEQGGWRLAPITAYSQOTRGLGCIIT 219
Qy 181 SLTGRDNQVGEVGVVSTATQSLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYND 240
Db 220 SLTGRDNQVGEVGVVSTATQSLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYND 279
Qy 241 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLSPPRVSYLKSS 300
Db 280 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLSPPRVSYLKSS 339
Qy 301 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db 340 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 380
```

```
RESULT 6
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
```

```
; CURRENT APPLICATION NUMBER: US/10/650,585
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-12
```

```
Query Match 100.0%; Score 1778; DB 16; Length 380;
Best Local Similarity 100.0%; Pred. No. 3,2e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VRGSDAIIILTCVAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACKLV 60
Db 40 VRGSDAIIILTCVAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACKLV 99
Qy 61 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMNHAQLRLAVAVEPIVFSMEVKIIT 120
Db 100 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMNHAQLRLAVAVEPIVFSMEVKIIT 159
Qy 121 WGADTAACGDIISGLPVASARGREILLGPADNFEQGGWRLAPITAYSQOTRGLGCIIT 180
Db 160 WGADTAACGDIISGLPVASARGREILLGPADNFEQGGWRLAPITAYSQOTRGLGCIIT 219
Qy 181 SLTGRDNQVGEVGVVSTATQSLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYND 240
Db 220 SLTGRDNQVGEVGVVSTATQSLATCVNGVCMVTFHAGSKTLAGPKGPITQMTYND 279
Qy 241 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLSPPRVSYLKSS 300
Db 280 QDLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLSPPRVSYLKSS 339
Qy 301 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 341
Db 340 GGPILCPGSHAVGIFRAAVCTRGVAKAVDFIPVESMETTMR 380
```

```
RESULT 7
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11
```

```
Query Match 100.0%; Score 1778; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 3,4e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VRGSDAIIILTCVAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACKLV 60
Db 53 VRGSDAIIILTCVAVHPELIFDITKLLAIFGPLMVLQAGITKVPYFVRAQGLIRACKLV 112
Qy 61 RKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMNHAQLRLAVAVEPIVFSMEVKIIT 120
```

```

Db      113 RKAAGHYVQMAFMKLAALGTYYDHLTPLODMAHAGRLDAVAEPIFSDMVEKITT 172
Qy      121 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
Db      173 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 232
Qy      181 SLTGRDNQVEGEVQVSTATOSFLATCVNGVCTWFHAGSKTLGAPKPIITQYTNVD 240
Db      233 SLTGRDNQVEGEVQVSTATOSFLATCVNGVCTWFHAGSKTLGAPKPIITQYTNVD 292
Qy      241 QDLVGMQAPPGASMTPTCTGSSDLYLVTRHADVIYVRRRGDSRGLSPRPVSYLKSS 300
Db      293 QDLVGMQAPPGASMTPTCTGSSDLYLVTRHADVIYVRRRGDSRGLSPRPVSYLKSS 352
Qy      301 GGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPVESMETTR 341
Db      353 GGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPVESMETTR 393

```

```

RESULT 8
US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

```

```

Query Match      100.0%; Score 1778; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 3.4e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 60
Db      53 VRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 112
Qy      61 RKAAGHYVQMAFMKLAALGTYYDHLTPLODMAHAGRLDAVAEPIFSDMVEKITT 120
Db      113 RKAAGHYVQMAFMKLAALGTYYDHLTPLODMAHAGRLDAVAEPIFSDMVEKITT 172
Qy      121 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
Db      173 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 232
Qy      181 SLTGRDNQVEGEVQVSTATOSFLATCVNGVCTWFHAGSKTLGAPKPIITQYTNVD 240
Db      233 SLTGRDNQVEGEVQVSTATOSFLATCVNGVCTWFHAGSKTLGAPKPIITQYTNVD 292
Qy      241 QDLVGMQAPPGASMTPTCTGSSDLYLVTRHADVIYVRRRGDSRGLSPRPVSYLKSS 300
Db      293 QDLVGMQAPPGASMTPTCTGSSDLYLVTRHADVIYVRRRGDSRGLSPRPVSYLKSS 352
Qy      301 GGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPVESMETTR 341
Db      353 GGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPVESMETTR 393

```

```

RESULT 9
US-10-017-736-2
; Sequence 2, Application US/10017736

```

```

; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2

```

```

Query Match      100.0%; Score 1778; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.6e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 60
Db      57 VRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 116
Qy      61 RKAAGHYVQMAFMKLAALGTYYDHLTPLODMAHAGRLDAVAEPIFSDMVEKITT 120
Db      117 RKAAGHYVQMAFMKLAALGTYYDHLTPLODMAHAGRLDAVAEPIFSDMVEKITT 176
Qy      121 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSOQTRGLGCIIT 236
Qy      181 SLTGRDNQVEGEVQVSTATOSFLATCVNGVCTWFHAGSKTLGAPKPIITQYTNVD 240
Db      237 SLTGRDNQVEGEVQVSTATOSFLATCVNGVCTWFHAGSKTLGAPKPIITQYTNVD 296
Qy      241 QDLVGMQAPPGASMTPTCTGSSDLYLVTRHADVIYVRRRGDSRGLSPRPVSYLKSS 300
Db      297 QDLVGMQAPPGASMTPTCTGSSDLYLVTRHADVIYVRRRGDSRGLSPRPVSYLKSS 356
Qy      301 GGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPVESMETTR 341
Db      357 GGPILCPSGHNAVGFRAAVCTRGVAKAVDFIPVESMETTR 397

```

```

RESULT 10
US-10-650-585-2
; Sequence 2, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-2

```

```

Query Match      100.0%; Score 1778; DB 16; Length 409;
Best Local Similarity 100.0%; Pred. No. 3.6e-174;
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 VRGGRDAIILITCAVHPELIFDITKLLAIFGPMVLQAGITKVPYFVRAQGLIRACMLV 60

```

```

Db      57 VRGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 116
Qy      61 RKAAGHYVQMAFKMLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPFIISDMEVKIIT 120
Db      117 RKAAGHYVQMAFKMLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPFIISDMEVKIIT 176
Qy      121 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 236
Qy      181 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFHGAGSKTLAGPKGITOMYTND 240
Db      237 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFHGAGSKTLAGPKGITOMYTND 296
Qy      241 QDLVGMQAPPGARSMTPTCTCGSSDLVYTRHADVIPIVRRRDSRGLSPRPVSYLKSS 300
Db      297 QDLVGMQAPPGARSMTPTCTCGSSDLVYTRHADVIPIVRRRDSRGLSPRPVSYLKSS 356
Qy      301 GGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
Db      357 GGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 397

```

```

RESULT 11
US-10-029-907-3
; Sequence 3, Application US/10029907
; Publication No. US20020142350A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/029, 907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257, 857
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
US-10-029-907-3

```

```

Query Match      95.7%; Score 1702; DB 13; Length 2201;
Best Local Similarity 94.1%; Pred. No. 2.6e-165;
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

```

Qy      1 VRGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 60
Db      57 VRGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 116
Qy      61 RKAAGHYVQMAFKMLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPFIISDMEVKIIT 120
Db      117 RKAAGHYVQMAFKMLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPFIISDMEVKIIT 176
Qy      121 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 236
Qy      181 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFHGAGSKTLAGPKGITOMYTND 240
Db      237 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFHGAGSKTLAGPKGITOMYTND 296
Qy      241 QDLVGMQAPPGARSMTPTCTCGSSDLVYTRHADVIPIVRRRDSRGLSPRPVSYLKSS 300

```

```

Db      297 QDLVGMQAPPGARSMTPTCTCGSSDLVYTRHADVIPIVRRRDSRGLSPRPVSYLKSS 356
Qy      301 GGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
Db      357 GGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 397

```

```

RESULT 12
US-10-309-561-3
; Sequence 3, Application US/10309561
; Publication No. US20030148348A1
; GENERAL INFORMATION:
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM
; FILE REFERENCE: 13/083
; CURRENT APPLICATION NUMBER: US/10/309, 561
; PRIOR FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US/10/029, 907
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257, 857
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: HCV
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 882
; OTHER INFORMATION: Xaa is Lys or Arg
; NAME/KEY: VARIANT
; LOCATION: 1489
; OTHER INFORMATION: Xaa is Leu
US-10-309-561-3

```

```

Query Match      95.7%; Score 1702; DB 14; Length 2201;
Best Local Similarity 94.1%; Pred. No. 2.6e-165;
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```

```

Qy      1 VRGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 60
Db      57 VRGRDAIILTCVAHPELIFDITKLLAIFGPMVLQAGITKVPFVRAQGLIRACMLV 116
Qy      61 RKAAGHYVQMAFKMLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPFIISDMEVKIIT 120
Db      117 RKAAGHYVQMAFKMLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPFIISDMEVKIIT 176
Qy      121 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 180
Db      177 WGAADTAACGDIISGLPVSARREIILGPADNFEQGWRLAPITAYSOQTRGLGCIIT 236
Qy      181 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFHGAGSKTLAGPKGITOMYTND 240
Db      237 SLTGRDNQVGEVGVVSTATQSFATCNGVCWTFHGAGSKTLAGPKGITOMYTND 296
Qy      241 QDLVGMQAPPGARSMTPTCTCGSSDLVYTRHADVIPIVRRRDSRGLSPRPVSYLKSS 300
Db      297 QDLVGMQAPPGARSMTPTCTCGSSDLVYTRHADVIPIVRRRDSRGLSPRPVSYLKSS 356
Qy      301 GGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 341
Db      357 GGFLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETTR 397

```

APPLICANT: De Francesco, Raffaele
APPLICANT: Migliaccio, Giovanni
APPLICANT: Paonessa, Giacomo
TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON
TITLE OF INVENTION: ENHANCED CELLS
FILE REFERENCE: IT00003P
CURRENT APPLICATION NUMBER: US/10/467,000
CURRENT FILING DATE: 2003-07-21
PRIOR APPLICATION NUMBER: PCT/EP02/00526
PRIOR FILING DATE: 2002-01-16
PRIOR APPLICATION NUMBER: 60/263,479
PRIOR FILING DATE: 2001-01-23
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3010
TYPE: PRT
ORGANISM: Con 1 HCV isolate nucleic acid
US-10-467-000-1

Query Match 95.7%; Score 1702; DB 12; Length 3010;
Best Local Similarity 94.1%; Pred. No. 4,1e-165;
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VGGSDAIIITCAVHPELIPDITKLLAIFGPMVLOAGITKPYFVRAQGLIRACMLV 60
DB 866 VGGSDAIIITCAVHPELIPDITKLLAIFGPMVLOAGITKPYFVRAQGLIRACMLV 925
QY 61 KKAAGHYVQMAFMKLAALITGYVDHLTPLOMAHAGLDLAAVEPVFSDMEVKIITW 120
DB 926 KKAAGHYVQMAFMKLAALITGYVDHLTPLOMAHAGLDLAAVEPVFSDMEVKIITW 985
QY 121 WGNADTAACGDIISGLPVASARRREIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 180
DB 986 WGNADTAACGDIISGLPVASARRREIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 1045
QY 181 SITGSDKNOVEGEVGVSTATQSFPLATCVNGCVCTVFGAGSKTLAPKGPITQMTTNDV 240
DB 1046 SITGSDKNOVEGEVGVSTATQSFPLATCVNGCVCTVFGAGSKTLAPKGPITQMTTNDV 1105
QY 241 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSRGSILSPRPVSYLKSSG 300
DB 1106 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSRGSILSPRPVSYLKSSG 1165
QY 301 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFVPSMETTW 341
DB 1166 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFVPSMETTW 1206

RESULT 14
US-09-919-901-4
Sequence 4, Application US/09919901
Publication No. US20030082518A1
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patlick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/919,901
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-02-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: :
US-09-919-901-4

Query Match 95.5%; Score 1698; DB 10; Length 1692;
Best Local Similarity 94.1%; Pred. No. 4.6e-165;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGRDAIILTCVHPELIPDITKLLAIFGPMVLOAGITKPYFVRAQGLIRACMLV 61
DB 146 RGRDAIILTCVHPELIPDITKLLAIFGPMVLOAGITKPYFVRAQGLIRACMLV 205
QY 62 KKAAGHYVQMAFMKLAALITGYVDHLTPLOMAHAGLDLAAVEPVFSDMEVKIITW 121
DB 206 KKAAGHYVQMAFMKLAALITGYVDHLTPLOMAHAGLDLAAVEPVFSDMEVKIITW 265
QY 122 GADTAACGDIISGLPVASARRREIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 181
DB 266 GADTAACGDIISGLPVASARRREIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 325
QY 182 LTGRPKNOVEGEVGVSTATQSFPLATCVNGCVCTVFGAGSKTLAPKGPITQMTTNDV 241
DB 326 LTGRPKNOVEGEVGVSTATQSFPLATCVNGCVCTVFGAGSKTLAPKGPITQMTTNDV 385
QY 242 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSRGSILSPRPVSYLKSSG 301
DB 386 DLVGMQAPPGARSMTPCTCGSSDLVLTTRADVIPIVRRRDSRGSILSPRPVSYLKSSG 445
QY 302 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFVPSMETTW 341
DB 446 GGPLCPGSHAVGIFRAAVCTRGVAKAVDPFVPSMETTW 495

RESULT 15
US-10-191-966-4
Sequence 4, Application US/10191966
Publication No. US20030175692A1
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patlick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/10/191,966
CURRENT FILING DATE: 2002-07-10
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-03-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: :
US-10-191-966-4

Query Match 95.5%; Score 1698; DB 14; Length 1692;
Best Local Similarity 94.1%; Pred. No. 4.6e-165;
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGRDAIILTCVHPELIPDITKLLAIFGPMVLOAGITKPYFVRAQGLIRACMLV 61
DB 146 RGRDAIILTCVHPELIPDITKLLAIFGPMVLOAGITKPYFVRAQGLIRACMLV 205
QY 62 KKAAGHYVQMAFMKLAALITGYVDHLTPLOMAHAGLDLAAVEPVFSDMEVKIITW 121
DB 206 KKAAGHYVQMAFMKLAALITGYVDHLTPLOMAHAGLDLAAVEPVFSDMEVKIITW 265
QY 122 GADTAACGDIISGLPVASARRREIILGPADNFEQGRLLAPITAYSOQTRGLGCIIT 181

Fri May 7 13:37:05 2004

us-10-650-585-14.rapb

Page 7

Db 266 GADTAACGDIILGLPVSARKEIILGPADSLIEGRWRLAPITVYSQOTRGLGCIITS 325
QY 182 LTGRDKXQVEGEVQVSTATQSFATCVCNGVCMVTFPGAGSKTLAGPKGPITOMYTNVDQ 241
Db 326 LTGRDKXQVEGEVQVSTATQSFATCVCNGVCMVTFPGAGSKTLAGPKGPITOMYTNVDQ 385
QY 242 DLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVI PVRRRGDSRGSLSPRPVSYLKGSSG 301
Db 386 DLVGMQAPPGARSLTPCTCGSSDLYLVTRHADVI PVRRRGDSRGSLSPRPVSYLKGSSG 445
QY 302 GPLLCPGSHAVGIFRAVCTGSAKAVDPIPVESMETTMR 341
Db 446 GPLLCPGSHAVGIFRAVCTGSAKAVDPIPVESMETTMR 485

Search completed: May 6, 2004, 09:43:22
Job time : 35.8831 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 39.7461 Seconds
(without alignments)
2075.771 Million cell updates/sec

Title: US-10-650-585-15
Perfect score: 1532
Sequence: 1 AAGGIRKCMVLRKAGSHYV.....RGVAKAVDFIPVESMTTWR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1532	100.0	292	5	ABG32188 HCV prote
2	1532	100.0	303	5	ABG32183 HCV prote
3	1532	100.0	334	5	ABG32182 HCV prote
4	1532	100.0	341	5	ABG32187 HCV prote
5	1532	100.0	352	5	ABG32186 HCV prote
6	1532	100.0	380	5	ABG32185 HCV prote
7	1532	100.0	393	5	ABG32184 HCV prote
8	1532	100.0	409	5	ABG32181 HCV prote
9	1523	99.4	303	5	ABG32191 HCV prote
10	1522	99.3	303	5	ABG32189 HCV prote
11	1513	98.8	301	5	ABG32190 HCV prote
12	1481	96.7	768	2	AAR40223 Recombina
13	1476	96.3	3010	5	AAE20477 HCV-S1 fu
14	1473	96.1	2201	5	ABG30601 Hepatitis
15	1473	96.1	2201	5	ABG30591 Hepatitis
16	1473	96.1	2201	5	ABG30600 Hepatitis
17	1473	96.1	2201	5	ABG30581 Hepatitis
18	1473	96.1	2201	5	ABG30593 Hepatitis
19	1473	96.1	2201	5	ABG30582 Hepatitis
20	1473	96.1	2201	5	ABG30580 Hepatitis
21	1473	96.1	2201	5	ABG30587 Hepatitis
22	1473	96.1	2201	5	ABG30599 Hepatitis
23	1473	96.1	2201	5	ABG30594 Hepatitis
24	1473	96.1	2201	5	ABG30598 Hepatitis
25	1473	96.1	2201	5	ABG30595 Hepatitis

ALIGNMENTS

25	1473	96.1	2307	3	AAY70064	AAY70064 Recombina
27	1473	96.1	3010	2	AAR68822	AAR68822 HCV prote
28	1473	96.1	3010	2	AAR62694	AAR62694 Partial H
29	1473	96.1	3010	5	ABG32458	ABG32458 Hepatitis
30	1473	96.1	3010	5	ABG32459	ABG32459 Hepatitis
31	1473	96.1	3010	5	ABG32451	ABG32451 Hepatitis
32	1473	96.1	3010	5	ABG32455	ABG32455 Hepatitis
33	1473	96.1	3010	5	ABG32457	ABG32457 Hepatitis
34	1473	96.1	3010	5	ABG32460	ABG32460 Hepatitis
35	1473	96.1	3010	5	ABG32461	ABG32461 Hepatitis
36	1473	96.1	3010	5	ABG32454	ABG32454 Hepatitis
37	1473	96.1	3011	2	ABG32456	ABG32456 Hepatitis
38	1470	96.0	344	2	AAR62695	AAR62695 HCV parti
39	1470	96.0	2201	5	ABG30586	ABG30586 Hepatitis
40	1470	96.0	2201	5	ABG30589	ABG30589 Hepatitis
41	1470	96.0	2201	5	ABG30583	ABG30583 Hepatitis
42	1470	96.0	2201	5	ABG30588	ABG30588 Hepatitis
43	1470	96.0	2307	3	AAY70065	AAY70065 Recombina
44	1469	95.9	2201	5	ABG30590	ABG30590 Hepatitis
45	1467	95.8	3010	2	AAR68864	AAR68864 Hepatitis

RESULT 1
ID ABG32188 standard; protein; 292 AA.

XX ABG32188;
XX 05-NOV-2002 (first entry)
XX HCV protease NS2/3 truncation mutant 915-1206.

KM HCV, enzyme; protease; NS2/3 (915-1206); hepatitis C virus infection;
KM Chronic liver disease; cirrhosis; end-stage liver disease; virulence;
KM hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;
KM chaotropic agent; mutant; mutein.

XX Hepatitis C virus.
OS Synthetic.

PN WO200248375-A2.

XX 20-JUN-2002.

PF 13-DEC-2001; 2001MO-CN001796.

XX 15-DEC-2000; 2000US-0256031P.

PA (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;
XX WPI, 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases
PT useful as therapeutic agents against hepatitis C virus, comprises full
PT length non-structural protease, or its truncation.

XX Claim 41; Page 63; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
XX length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
XX to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
XX residue amino acid 810 to 906, or having a minimal amino acid sequence
XX from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
XX NS2/3 protease. Also included are (1) a composition (c) comprising an
XX isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
XX its truncation or a mutated sequence, where the protease is in a solution
XX comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
XX to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC; and (6) screening a potential inhibitor of M3 in the presence of, or
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 915-1206
 CC (numbered relative to the full length NS2/3 protein)

CC Sequence 292 AA;

Query Match 100.0%; Score 1532; DB 5; Length 292;
 Best Local Similarity 100.0%; Pred. No. 1.2e-140;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAGHYVQMAFMKLAALTGTYVYDHLPLQDMHAGRLDAVAVEPV 60
 DB 1 AAGIIRACMLVRKAGHYVQMAFMKLAALTGTYVYDHLPLQDMHAGRLDAVAVEPV 60
 QY 61 IFSDEVKIITWGDADTAACGDIISGLPVSARGREIILGPDNFEQGMRLAIPITAYSQ 120
 DB 61 IFSDEVKIITWGDADTAACGDIISGLPVSARGREIILGPDNFEQGMRLAIPITAYSQ 120
 QY 121 QTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAIPK 180
 DB 121 QTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAIPK 180
 QY 181 GPIQMTNTVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVPVRRRGDSRGSLLS 240
 DB 181 GPIQMTNTVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVPVRRRGDSRGSLLS 240
 QY 241 PRPVSYLKSSGGGPLLCPSGHNVGIFRAVCTRGVAKAVDPFVPSMETTMR 292
 DB 241 PRPVSYLKSSGGGPLLCPSGHNVGIFRAVCTRGVAKAVDPFVPSMETTMR 292

RESULT 2
 ABG32198
 ID ABG32183 standard; protein; 303 AA.

AC ABG32183;

DT 05-NOV-2002 (first entry)

DE HCV protease NS2/3 truncation mutant 904-1206.

KW HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
 KW chronic liver disease; cirrhosis; end-stage liver disease; viremia;
 KW hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KW chaotropic agent; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

XX WO200248375-A2.

PD 20-JUN-2002.

XX

PF 13-DEC-2001; 2001WO-CA001796.
 XX
 XX 15-DEC-2000; 2000US-0256031P.
 PR
 PA (BOH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 PT Thibeault D, Lamarre D, Maurice R, Paloce L, Pause A,
 PI WPI; 2002-599511/64.
 DR
 PT Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.

Claim 39; Page 58-59; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 to also as NS2/3 (910-1206)) or its truncation, having as its N-terminal
 residue amino acid 810 to 906, or having a minimal amino acid sequence
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 904-1206
 CC (numbered relative to the full length NS2/3 protein)

CC Sequence 303 AA;

Query Match 100.0%; Score 1532; DB 5; Length 303;
 Best Local Similarity 100.0%; Pred. No. 1.2e-140;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAGHYVQMAFMKLAALTGTYVYDHLPLQDMHAGRLDAVAVEPV 60
 DB 1 AAGIIRACMLVRKAGHYVQMAFMKLAALTGTYVYDHLPLQDMHAGRLDAVAVEPV 60
 QY 61 IFSDEVKIITWGDADTAACGDIISGLPVSARGREIILGPDNFEQGMRLAIPITAYSQ 120
 DB 61 IFSDEVKIITWGDADTAACGDIISGLPVSARGREIILGPDNFEQGMRLAIPITAYSQ 120
 QY 121 QTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAIPK 180
 DB 121 QTRGLGCIITSLTGRDKNOVEGEVQVSTATQSFATCVNGVCMVTFHAGSKTLAIPK 180
 QY 181 GPIQMTNTVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVPVRRRGDSRGSLLS 240
 DB 181 GPIQMTNTVDQDLVGMQAPPGASMTPTCGSSDLYLVTRHADVPVRRRGDSRGSLLS 240
 QY 241 PRPVSYLKSSGGGPLLCPSGHNVGIFRAVCTRGVAKAVDPFVPSMETTMR 292
 DB 241 PRPVSYLKSSGGGPLLCPSGHNVGIFRAVCTRGVAKAVDPFVPSMETTMR 292

Db	252	PREVSTLKGSGGAPLCPGSHAVGIFRAAVCRGAKAVDFIPVESHETTVR	303
RESULT 3			
ABG32182			
ID	ABG32182	standard; protein; 334 AA.	
XX			
AC	ABG32182;		
XX			
DT	05-NOV-2002	(first entry)	
XX			
DE	HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.		
XX			
KW	HCV; enzyme; protease; NS2/3 (910-1206); hepatitis C virus infection;		
KW	chronic liver disease; cirrhosis; end-stage liver disease; viraemia;		
KW	hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO;		
XX	chaotropic agent; 4K-6H (904-1206)st-4K; mutant; mutein.		
XX			
OS	Hepatitis C virus.		
OS	Synthetic.		
FX			
FX	Key	Location/Qualifiers	
FT	Peptide	1..15	
FT		/note="4-Lys/His tag"	
FT	Protein	16..302	
FT		/note="Truncated NS2/3 protease"	
FT	Peptide	319..334	
FT		/note="Streptavidin/4-Lys tag"	
FN	WC0200248375-A2.		
PN			
PD	20-JUN-2002.		
XX			
PF	13-DEC-2001; 2001MO-CA001796.		
XX			
XX	15-DEC-2000; 2000US-0256031P.		
XX			
XX	(BOEH) BOEHRINGER INGELHEIM CANADA LTD.		
XX			
PA	Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;		
XX			
PI	WPI; 2002-5595511/64.		
DR	N-PSDS; ABR90407.		
XX			
XX			
PT	Novel polypeptide for screening inhibitors of non-structural proteases		
PT	useful as therapeutic agents against hepatitis C virus, comprises full		
PT	length non-structural protease, or its truncation.		
PS	Claim 39; Fig 9B; 67pp; English.		
XX			
XX			
CC	The invention relates to an isolated polypeptide consisting of a full-		
CC	length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred		
CC	to also as NS2/3 (910-1206)), or its truncation, having as its N-terminal		
CC	residue amino acid 810 to 906, or having a minimal amino acid sequence		
CC	from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length		
CC	NS2/3 protease. Also included are (1) a composition (C) comprising an		
CC	isolated HCV NS2/3 protease selected from full length NS2/3 protease, or		
CC	its truncation or a mutated sequence, where the protease is in a solution		
CC	comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)		
CC	to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide		
CC	appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3		
CC	protease, involving isolating the protease in the presence of a		
CC	chaotropic agent, refolding the isolated protease by contacting it with a		
CC	reducing agent, and LDAO in the presence of reduced concentration of the		
CC	chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3		
CC	protease, involving diluting refolded inactive NS2/3 protease in a medium		
CC	containing an activation detergent to induce auto-cleavage of the NS2/3		
CC	protease; (5) measuring (M3) the auto-cleavage activity of NS2/3		
CC	protease, involving incubating the active NS2/3 protease produced by M2		
CC	for sufficient time to induce auto-cleavage of NS2/3 protease and produce		
CC	cleavage products or their fragments, and measuring the presence or		
CC	absence of uncleaved NS2/3 protease, cleavage products or their fragments		

Query Match	100.0%	Score 1532	DB 5	Length 334
Best Local Similarity	100.0%	Pred. No. 1,4e-140		
Matches 292	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Db	1	AGGLIACMLVKKRAGGHHVQVAFPKLALNGTYVYDHLTPLODMAHGLRDLAVAEPV	60	
Db	27	AGGLIACMLVKKRAGGHHVQVAFPKLALNGTYVYDHLTPLODMAHGLRDLAVAEPV	86	
Qy	61	IFSDMEVKIITWGDADTACGDIISGLPVSRGRRIILGPADNFEQGMRLAPITYSQ	120	
Db	87	IFSDMEVKIITWGDADTACGDIISGLPVSRGRRIILGPADNFEQGMRLAPITYSQ	146	
Qy	121	QTFGLIGCIITSLTQBDKNQVGEVQVSTATQSFATCVNGVQMTVFHAGSKYTLAGPK	180	
Db	147	QTFGLIGCIITSLTQBDKNQVGEVQVSTATQSFATCVNGVQMTVFHAGSKYTLAGPK	206	
Qy	181	GPITQMTVTVNDQPLVGMQAPFGARSMTPTCTGSSDPLVYTRHADVIPIRRRGDSRGSLLS	240	
Db	207	GPITQMTVTVNDQPLVGMQAPFGARSMTPTCTGSSDPLVYTRHADVIPIRRRGDSRGSLLS	266	
Qy	241	PRVSTVTKSSGGPILCPSGHNVGIFPRAVCTRGVAKAVDPIPVESVETTW	292	
Db	267	PRVSTVTKSSGGPILCPSGHNVGIFPRAVCTRGVAKAVDPIPVESVETTW	318	
RESULT 4				
ABG32187	ID	ABG32187	standard; protein; 341 AA.	
XX	XX	ABG32187;		
XX	XX	05-NOV-2002 (first entry)		
XX	XX	HCV protease NS2/3 truncation mutant 866-1206.		
XX	XX	HCV, enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection;		
XX	XX	chronic liver disease; cirrhosis; end-stage liver disease; virucide;		
XX	XX	hepatotropic; antiinflammatory; lauryldiethylenamine oxide; IDA0;		
XX	XX	chaotropic agent; mutant; mutein.		
XX	XX	Hepatitis C virus.		
XX	XX	Synthetic.		
XX	XX	W0200248375-A2.		
XX	XX	20-JUN-2002.		
XX	XX	13-DEC-2001; 2001MO-CA001796.		
XX	XX	15-DEC-2000; 2000US-0256031P.		
XX	XX	(BOEHRINGER INGELHEIM CANADA LTD.		
XX	XX	Thibault D, Lamarre D, Maurice R, Pilote L, Pause A;		
XX	XX	WPI; 2002-559511/64.		
XX	XX	Novel polypeptide for screening inhibitors of non-structural proteases		

PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.

PS Claim 41; Page 62-63; 67pp; English.

XX
 CC The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 866-1206
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 341 AA;

Query Match 100.0%; Score 1532; DB 5; Length 341;
 Best Local Similarity 100.0%; Pred. No. 1.5e-140; Indels 0; Gaps 0;
 Matches 292; Conservative 0; Mismatches 0;

QY 1 AAGLRACMLVRKAAAGHYVQMAFMKLAALGTIVYDHLTPLDWMAGRLDAVAPEV 60
 DB 50 AAGLRACMLVRKAAAGHYVQMAFMKLAALGTIVYDHLTPLDWMAGRLDAVAPEV 109
 QY 61 IFSMVEKLTITWGDPTAACGDTISGLFVSARRGREILGADNPEGGWMLLPIRAYSQ 120
 DB 110 IFSMVEKLTITWGDPTAACGDTISGLFVSARRGREILGADNPEGGWMLLPIRAYSQ 169
 QY 121 QTRGLGCIITSLGRDNQVEGEVQVSTATQSFLLATCVNGVCMVFEHAGSKTLAEPK 180
 DB 170 QTRGLGCIITSLGRDNQVEGEVQVSTATQSFLLATCVNGVCMVFEHAGSKTLAEPK 229
 QY 181 GPITQMTNVDDIVGQAPPGARSMTPTCGSSDLYLVTRHADYIFVRRRGDSRGLLS 240
 DB 230 GPITQMTNVDDIVGQAPPGARSMTPTCGSSDLYLVTRHADYIFVRRRGDSRGLLS 289
 QY 241 PRPVSYLKGGSGGFLPCPSGAVGIFRAAVCTRGVAFAVPIFVESMETMR 292
 DB 290 PRPVSYLKGGSGGFLPCPSGAVGIFRAAVCTRGVAFAVPIFVESMETMR 341

RESULT 5

ABG32186 ID ABG32186 standard; protein; 352 AA.

AC ABG32186;

XX 05-NOV-2002 (first entry)

XX HCV protease NS2/3 truncation mutant 855-1206.

XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;

XX chronic liver disease; cirrhosis; end-stage liver disease; viraemia;

XX hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;

XX chaotropic agent; mutant; mutain.

XX Hepatitis C virus.

XX Synthetic.

XX WO200248375-A2.

XX 20-JUN-2002.

XX 13-DEC-2001; 2001WO-CA001796.

XX 15-DEC-2000; 2000US-0256031P.

XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.

XX Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;

XX WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases

XX useful as therapeutic agents against hepatitis C virus, comprises full

XX length non-structural protease, or its truncation.

XX Claim 41; Page 61-62; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 855-1206
 CC (numbered relative to the full length NS2/3 protein)

XX Sequence 352 AA;

Query Match 100.0%; Score 1532; DB 5; Length 352;
 Best Local Similarity 100.0%; Pred. No. 1.5e-140; Indels 0; Gaps 0;
 Matches 292; Conservative 0; Mismatches 0;

QY 1 AAGLRACMLVRKAAAGHYVQMAFMKLAALGTIVYDHLTPLDWMAGRLDAVAPEV 60

Db 61 AAGLIRACMLVKAAGHYVQAFMKLALITGYVYDHLPLQDMAHAGRLAVAVEPV 120
 QY 61 IFSDEVEKIIITWADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAPITAVSQ 120
 Db 121 IFSDEVEKIIITWADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAPITAVSQ 180
 QY 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSFLATCNGVCMTVPFGAGSKTLAEPK 180
 Db 181 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSFLATCNGVCMTVPFGAGSKTLAEPK 240
 QY 181 GPITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRDSGSLIS 240
 Db 241 GPITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRDSGSLIS 300
 QY 241 PRPVSYLKSSSGGPLLCPSGHANGIFRAAVCTRGVAKADVPIVSEMETTMR 292
 Db 301 PRPVSYLKSSSGGPLLCPSGHANGIFRAAVCTRGVAKADVPIVSEMETTMR 352

RESULT 6

ABG32185
 ID ABG32185 standard; protein; 380 AA.
 XX
 AC ABG32185;
 XX
 DT 05-NOV-2002 (first entry)
 XX

HCV protease NS2/3 truncation mutant 827-1206.
 DE

XX HCV; enzyme; protease; NS2/3 (827-1206); hepatitis C virus infection;
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 KM hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KM chaotropic agent; mutant; muten.
 XX

OS Hepatitis C virus.
 OS Synthetic.
 XX

WO200248375-A2.
 XX

20-JUN-2002.
 PD

13-DEC-2001; 2001WO-CA001796.
 PF

15-DEC-2000; 2000US-0256031P.
 PR

(BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 PA

Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 PI

WPI; 2002-599511/64.
 XX

Novel polypeptide for screening inhibitors of non-structural proteases
 useful as therapeutic agents against hepatitis C virus, comprises full
 length non-structural protease, or its truncation.
 PT

Claim 41; Page 60-61; 67pp; English.
 PS

The invention relates to an isolated polypeptide consisting of a full-
 length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 residue amino acid 810 to 906, or having a minimal amino acid sequence
 from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 NS2/3 protease. Also included are (1) a composition (C) comprising an
 isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 its truncation or a mutated sequence, where the protease is in a solution
 comprising a sufficient concentration of lauryldiethylamine oxide (LDAO),
 to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 appearing as ABG32185; (3) producing (M1) a refolded, inactive HCV NS2/3
 protease, involving isolating the protease in the presence of a
 chaotropic agent, refolding the isolated protease by contacting it with a
 reducing agent, and LDAO in the presence of reduced concentration of the
 chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 protease, involving diluting refolded inactive NS2/3 protease in a medium

containing an activation detergent to induce auto-cleavage of the NS2/3
 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 protease, involving incubating the active NS2/3 protease produced by M2
 for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 cleavage products or their fragments, and measuring the presence or
 absence of uncleaved NS2/3 protease, cleavage products or their fragments
 ; and (6) screening a potential inhibitor of auto-cleavage activity of an
 active NS2/3 protease, involving carrying out M3 in the presence of, or
 absence of the potential inhibitor, comparing the amount of uncleaved
 NS2/3 protease, cleavage products or their fragments. The protease is
 useful for detailed biochemical characterization of the enzymes and in
 the development of in vitro assays for screening novel inhibitors of
 NS2/3 protease which are useful as therapeutic agents against HCV
 infection (which causes chronic liver disease, cirrhosis and end-stage
 liver disease. M1 is useful for high level production of protease. The
 present sequence represents the NS2/3 truncation mutant 827-1206
 (numbered relative to the full length NS2/3 protein)

Sequence 380 AA;
 XX

Query Match 100.0%; Score 1532; DB 5; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.7e-140;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVKAAGHYVQAFMKLALITGYVYDHLPLQDMAHAGRLAVAVEPV 60
 Db 89 AAGLIRACMLVKAAGHYVQAFMKLALITGYVYDHLPLQDMAHAGRLAVAVEPV 148

QY 61 IFSDEVEKIIITWADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAPITAVSQ 120
 Db 149 IFSDEVEKIIITWADTAACGDIISGLPVSARRGREILLGPADNFEQGWRLAPITAVSQ 208

QY 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSFLATCNGVCMTVPFGAGSKTLAEPK 180
 Db 209 QTRGLGCIITSLTGRDKNQVEGEVQVSTATOSFLATCNGVCMTVPFGAGSKTLAEPK 268

QY 181 GPITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRDSGSLIS 240
 Db 269 GPITQWNTVNDODLVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRRDSGSLIS 328

QY 241 PRPVSYLKSSSGGPLLCPSGHANGIFRAAVCTRGVAKADVPIVSEMETTMR 292
 Db 329 PRPVSYLKSSSGGPLLCPSGHANGIFRAAVCTRGVAKADVPIVSEMETTMR 380

RESULT 7
 ABG32184
 ID ABG32184 standard; protein; 393 AA.
 XX

ABG32184;
 AC

05-NOV-2002 (first entry)
 DT

HCV protease NS2/3 truncation mutant 815-1206.
 DE

XX HCV; enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection;
 KM chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
 KM hepatocytic; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KM chaotropic agent; mutant; muten.
 XX

OS Hepatitis C virus.
 OS Synthetic.
 XX

WO200248375-A2.
 PN

20-JUN-2002.
 PD

13-DEC-2001; 2001WO-CA001796.
 PF

15-DEC-2000; 2000US-0256031P.
 PR

(BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 PA

PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 XX
 XX WPI; 2002-599511/64.
 DR
 XX Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.
 XX
 XX Claim 41; Page 59-60; 67pp; English.
 XX
 CC The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation mutant 815-1206
 CC (numbered relative to the full length NS2/3 protein)
 XX
 XX Sequence 393 AA:
 SQ
 Query Match 100.0%; Score 1532; DB 5; Length 393;
 Best Local Similarity 100.0%; Pred. No. 1,8e-140;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGGIRACMLVRKAGSHYQVMAFMKLAALTGTYYVDHLPLDMDMAHGRDLAAVVEPV 60
 DB 102 AAGGIRACMLVRKAGSHYQVMAFMKLAALTGTYYVDHLPLDMDMAHGRDLAAVVEPV 161
 QY 61 IFSDMEVYKITTGWADTAACDIISGLPVSARRGREILIGPADNFEQGMRLAPITAYSQ 120
 DB 162 IFSDMEVYKITTGWADTAACDIISGLPVSARRGREILIGPADNFEQGMRLAPITAYSQ 221
 QY 121 QTRGLGACITTSLTGRPNKQGEVGVSVATQSPFATCNGVCMVFHGAGSKTLTAGKR 180
 DB 222 QTRGLGACITTSLTGRPNKQGEVGVSVATQSPFATCNGVCMVFHGAGSKTLTAGKR 281
 QY 181 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRADVIVARRGDSRSLLS 240
 DB 282 GPITQMTNVNDQDLVGMQAPPGARSMTPTCGSSDLYLVTRADVIVARRGDSRSLLS 341
 QY 241 PRVSVYTKGSSGGPILCPSGHAYGIFRAAVCTRGVAKAVDFIVESMETTMR 292
 DB 342 PRVSVYTKGSSGGPILCPSGHAYGIFRAAVCTRGVAKAVDFIVESMETTMR 393

ID ABG32181 standard; protein; 409 AA.
 XX
 AC ABG32181;
 XX
 DT 05-NOV-2002 (first entry)
 XX
 XX HCV protease NS2/3 (810-1206).
 DE
 XX
 XX HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection;
 XX chronic liver disease; cirrhosis; end-stage liver disease; virocidic;
 XX hepatotropic; antinflammatory; lauryldiethylamine oxide; LDAO;
 XX chaotropic agent; mutant; mutain.
 XX
 OS Hepatitis C virus.
 OS Synthetic.
 FH Key location/Qualifiers
 FT Peptide 398..409
 FT /note="Streptavidin tag"
 XX
 XX WO200248375-A2.
 XX
 XX 20-JUN-2002.
 XX
 XX 13-DEC-2001; 2001WO-CR001796.
 XX
 XX 15-DEC-2000; 2000US-0256031P.
 XX
 XX (BOER) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 XX WPI; 2002-599511/64.
 DR N-PSDB; ABK90406.
 XX
 PT Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.
 XX
 XX Claim 42; Fig 1b; 67pp; English.
 XX
 CC The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 (810-1206) protein, which has a C-

CC	to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
CC	residue amino acid 910 to 906, or having a minimal amino acid sequence
CC	from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
CC	NS2/3 protease. Also included are (1) a composition (C) comprising an
CC	isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
CC	its truncation or a mutated sequence, where the protease is in a solution
CC	comprising a sufficient concentration of lauryldimethylamine oxide (LDAO)
CC	to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
CC	appearing as ABG32189; (3) producing (M1) a refolded, inactive HCV NS2/3
CC	protease, involving isolating the protease in the presence of a
CC	chaotropic agent, refolding the isolated protease by contacting it with a
CC	reducing agent, and LDAO in the presence of reduced concentration of
CC	chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
CC	protease, involving diluting refolded inactive NS2/3 protease in a medium
CC	containing an activation detergent to induce auto-cleavage of the NS2/3
CC	protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
CC	protease, involving incubating the active NS2/3 protease produced by M2
CC	for sufficient time to induce auto-cleavage of NS2/3 protease and produce
CC	cleavage products or their fragments, and measuring the presence or
CC	absence of uncleaved NS2/3 protease, cleavage products or their fragments
CC	; and (6) screening a potential inhibitor of auto-cleavage activity of an
CC	active NS2/3 protease, involving carrying out M3 in the presence of, or
CC	absence of the potential inhibitor, comparing the amount of uncleaved
CC	NS2/3 protease, cleavage products or their fragments. The protease is
CC	useful for detailed biochemical characterisation of the enzymes and in
CC	the development of in vitro assays for screening novel inhibitors of
CC	NS2/3 protease which are useful as therapeutic agents against HCV
CC	infection (which causes chronic liver disease, cirrhosis and end-stage
CC	liver disease. M1 is useful for high level production of protease. The
CC	present sequence represents the NS2/3 truncation 904-1206 mutant
CC	Cys993Ala (numbered relative to the full length NS2/3 protein) a mutant
CC	devoid of autocatalytic activity
XX	
XX	Sequence 303 AA;
SQ	
Query Match	99.4%; Score 1523; DB 5; Length 303;
Best Local Similarity	99.7%; Pred. No. 9.3e-140;
Matches 221; Conservative	0; Mismatches 1; Indels 0; Gaps 0
QY	1 AAGLRACMLVKAAGHYVMAEMKLAALTGTYYVDHLTPLODMHAAGRLDAVAPEV 60
DB	12 AAGLRACMLVKAAGHYVMAEMKLAALTGTYYVDHLTPLODMHAAGRLDAVAPEV 71
QY	61 IFSDEVKITTMGAATACGIIISGLPVSARGSEILGGADNFEQGRIAPITAVSQ 120
DB	72 IFSDEVKITTMGAATTAAGDIIISGLPVSARGSEILGGADNFEQGRLLAPIAVSQ 131
QY	121 QTRGLLGCIITISLTNRDNQVEGEVOVASTAQTOSFLATCVNGVCWTFPHGASXTLAGPK 180
DB	132 QTRGLLGCIITISLTNRDNQVEGEVOVASTAQTOSFLATCVNGVCWTFPHGASXTLAGPK 191
QY	181 GPITQNTTNDODLVNGMPARGASMPPTCGSSDLVTVRHADVIPIVRPRGSRSLLS 240
DB	192 GPITQNTTNDODLVNGMPARGASMPPTCGSSDLVTVRHADVIPVRRRDSRSLLS 251
QY	241 PREVSYLTKSSGAPLLCPSGHAVIGIFRAVACTRSVAANADVPIPVESMETMR 292
DB	252 PRVSVTLKSSGAPLLCPSGHAVIGIFRAVACTRSVAANADVPIPVESMETMR 303
RESULT 10	
ID	ABG32189 standard; protein; 303 AA.
XX	ABG32189;
AC	
XX	05-NOV-2002 (first entry)
DT	
XX	
DE	HCV protease NS2/3 truncation 904-1206/His952Ala.
XX	
KW	HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
KX	chronic liver disease; cirrhosis; end-stage liver disease; viraemia;
KM	hepatotropic; antiinflammatory; lauryldimethylamine oxide; LDAO;

KM chaotropic agent; mutant; mutein.
 XX Hepatitis C virus.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 49 /note="Wild-type His substituted by Ala"
 FT
 XX WO200248375-A2.
 XX
 XX 20-JUN-2002.
 XX
 XX 13-DEC-2001; 2001WO-CA001796.
 XX
 XX 15-DEC-2000; 2000US-0256031P.
 XX
 XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 XX WPI; 2002-599511/64.
 XX
 XX Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.
 XX
 XX Example 7; Fig 8; 67pp; English.
 XX
 XX The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a
 CC reducing agent, and LDAO in the presence of reduced concentration of the
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3
 CC protease, involving inducing refolded inactive NS2/3 protease in a medium
 CC containing an activation detergent to induce auto-cleavage of the NS2/3
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3
 CC protease, involving incubating the active NS2/3 protease produced by M2
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce
 CC cleavage products or their fragments, and measuring the presence or
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments
 CC; and (6) screening a potential inhibitor of auto-cleavage activity of an
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or
 CC absence of the potential inhibitor, comparing the amount of uncleaved
 CC NS2/3 protease, cleavage products or their fragments. The protease is
 CC useful for detailed biochemical characterisation of the enzymes and in
 CC the development of in vitro assays for screening novel inhibitors of
 CC NS2/3 protease which are useful as therapeutic agents against HCV
 CC infection (which causes chronic liver disease, cirrhosis and end-stage
 CC liver disease. M1 is useful for high level production of protease. The
 CC present sequence represents the NS2/3 truncation 904-1206 mutant
 CC H1955A1a (numbered relative to the full length NS2/3 protein) a mutant
 CC devoid of autocatalytic activity
 XX
 XX Sequence 303 AA.
 XX
 XX Query Match 99.3%; Score 1522; DB 5; Length 303;
 XX Best Local Similarity 99.7%; Pired. No. 1.2e-139;
 XX Matches 291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 XX
 XX 1 AAGTAAACCTVRAAGHYVQMAFMALATLTGYVDPHTLPIDVNHAGLRLAVAVPEV 60
 XX 12 AGGIRACVLRAPAGGHVQMAFMALATLTGYVDPHTLPIDVNHAGLRLAVAVPEV 71

QY 61 IFSDMEVKITTWGADTAACGDIISGLPVASARGREILIGPADNPEGQGMALAPITANSQ 120
 DB 72 IFSDMEVKITTWGADTAACGDIISGLPVASARGREILIGPADNPEGQGMALAPITANSQ 131
 QY 121 QTRGLGCIITSLTGRDNQVGEVQVSTQSPFLATCNAGVCTVHAGSKTLGAPK 180
 DB 132 QTRGLGCIITSLTGRDNQVGEVQVSTQSPFLATCNAGVCTVHAGSKTLGAPK 191
 QY 181 GPITQMTNVDDLVGWAAPPGARSMPTCTGSSDLYVTREADVIPVRRGDSRGSLIS 240
 DB 192 GPITQMTNVDDLVGWAAPPGARSMPTCTGSSDLYVTREADVIPVRRGDSRGSLIS 251
 QY 241 PRPVSYLKGSGGGLPCPSGHAIVGFRAAVCTRGVAKAVDPIPVESMETTKR 292
 DB 252 PRPVSYLKGSGGGLPCPSGHAIVGFRAAVCTRGVAKAVDPIPVESMETTKR 303
 RESULT 11
 ABG32190
 ID ABG32190 standard; protein; 301 AA.
 XX
 XX AC ABG32190;
 XX
 XX 05-NOV-2002 (first entry)
 XX
 XX HCV protease NS2/3 truncation 904-1206/deltaLeu1026-Ala1027.
 XX
 XX HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection;
 KM chronic liver disease; cirrhosis; end-stage liver disease; virocidic;
 KM hepatocarcinoma; antiinflammatory; lauryldiethylamine oxide; LDAO;
 KM chaotropic agent; mutant; mutein.
 OS
 OS Hepatitis C virus.
 OS Synthetic.
 FT Misc-difference 122.123
 FT /note="Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"
 XX
 XX WO200248375-A2.
 XX
 XX 20-JUN-2002.
 XX
 XX 13-DEC-2001; 2001WO-CA001796.
 XX
 XX 15-DEC-2000; 2000US-0256031P.
 XX
 XX (BOEH) BOEHRINGER INGELHEIM CANADA LTD.
 XX
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;
 XX WPI; 2002-599511/64.
 XX
 XX Novel polypeptide for screening inhibitors of non-structural proteases
 PT useful as therapeutic agents against hepatitis C virus, comprises full
 PT length non-structural protease, or its truncation.
 XX
 XX Example 7; Page 64-65; 67pp; English.
 XX
 XX The invention relates to an isolated polypeptide consisting of a full-
 CC length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal
 CC residue amino acid 810 to 906, or having a minimal amino acid sequence
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or
 CC its truncation or a mutated sequence, where the protease is in a solution
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO)
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide
 CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3
 CC protease, involving isolating the protease in the presence of a
 CC chaotropic agent, refolding the isolated protease by contacting it with a

reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant delta1026-1027 (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity

Sequence 301 AA:

Query Match 98.8%; Score 1513; DB 5; Length 301;
Best Local Similarity 99.3%; Pred. No. 8.7e-139;
Matches 290; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

1 AAGLRACMLVRKAGHYVQAFMKLAALTGTYYVDHLPLQDMAHAGLDLAVAVEPV 60
12 AAGLRACMLVRKAGHYVQAFMKLAALTGTYYVDHLPLQDMAHAGLDLAVAVEPV 71
61 IPSDMEVKITITWGDATACGDIISGLPVSARREKILLGPADNFEQGMRLAPITAYSQ 120
72 IFSMEVKIITWGDATACGDIISGLPVSARREKILLGPADNFEQGMRL--PITAYSQ 129
121 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGAGSKTLAEPK 180
130 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGAGSKTLAEPK 189
181 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGLIS 240
190 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGLIS 249
241 PRPVSYLKSGSGGGLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
250 PRPVSYLKSGSGGGLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 301

RESULT 12

AAR40223 standard; protein; 768 AA.

AAR40223;

21-FEB-1994 (first entry)

Recombinant hepatitis C virus genomic protein.

Hepatitis C virus; screening; antiviral drugs.

Hepatitis C virus.

Hepatitis C virus.

Key Location/Qualifiers

Misc-difference 10 /note= "UUA encodes Ile."

Misc-difference 81..82 /note= "Nucleotide sequence encodes another Gly"

Duplication 528..768 /note= "Duplication of 241 amino acids at start of protein sequence."

Misc-difference 537 /note= "UUA encodes Ile."

JP05192160-A.
03-AUG-1993.
20-JAN-1992; 92JP-00028833.
20-JAN-1992; 92JP-00028833.
(BANYU PHARM CO LTD.
N-PSDB; AA048215.
Hepatitis C virus genomic RNA, cDNA and polypeptide - used for screening hepatitis C virus-specific antiviral drugs.
Claim 9, Page 4-6, 14pp; Japanese.

The protein is useful in the screening of HCV-specific antiviral drugs.

HCV cDNA was cloned from plasma. Plasmids pSR3241 and pSR2541 were prepared using the cDNA and plasmid pSR3241 was used to transform a COS-1 cell

Sequence 768 AA:

Query Match 96.7%; Score 1481; DB 2; Length 768;
Best Local Similarity 95.2%; Pred. No. 4.3e-135;
Matches 278; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

1 AAGLRACMLVRKAGHYVQAFMKLAALTGTYYVDHLPLQDMAHAGLDLAVAVEPV 60
176 AAGLRACMLVRKAGHYVQAFMKLAALTGTYYVDHLPLQDMAHAGLDLAVAVEPV 235
61 IPSDMEVKITITWGDATACGDIISGLPVSARREKILLGPADNFEQGMRLAPITAYSQ 120
72 IFSMEVKIITWGDATACGDIISGLPVSARREKILLGPADNFEQGMRL--PITAYSQ 295
121 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGAGSKTLAEPK 180
129 QTRGLGCIITSLTGRDKNQEVEGVVSTATQSFATCVNGVCMTVPHGAGSKTLAEPK 355
181 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGLIS 240
356 GPITQYTNVDQDLVGMQAPPGARSMTPTCGSSDLVLTVRHADVIPIVRRGDSRGLIS 415
241 PRPVSYLKSGSGGGLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
416 PRPVSYLKSGSGGGLCPGSHAVGIFRAAVCTRGVAKAVDFIPVSMETMR 467

RESULT 13

AAE20477 standard; protein; 3010 AA.

AAE20477;

01-JUL-2002 (first entry)

HCV-S1 full-length polypeptide.

Nucleic acid construct; expression cassette; non-coding region; NCR;

untranslated region; UTR; anti-viral drug; drug resistance; HCV-S1;

Hepatitis C virus.

Hepatitis C virus.

WO200208447-A2.

31-JAN-2002.

20-JUL-2001; 2001WO-IL000669.

24-JUL-2000; 2000US-0220248P.

(MOLE-) INST MOLECULAR & CELL BIOLOGY.
(EHRICH) EHRICH G.

Tan YH, Lim SG, Lim SP, Hong WJ;

WPI: 2002-280605/32.

N-BSD; MAD33038.

Novel nucleic acid construct useful for detecting the presence of RNA virus, comprises an expression cassette and a promoter operably linked to expression cassette for minus strand RNA transcription of the cassette.

Example 1; Page 70-81; 81pp; English.

The invention relates to nucleic acid construct which comprises an expression cassette including a first polynucleotide region including a 5' non-coding region (NCR) sequence of an RNA virus and at least an N-terminal portion of a coding sequence of an RNA virus, a second polynucleotide region including a 3' untranslated region (UTR) sequence of the RNA virus and at least a C-terminal portion of a coding sequence of the virus and a third polynucleotide region encoding a reporter molecule, flanked by first and second polynucleotide regions; and a promoter sequence being operatively linked to expression cassette in a manner so as to enable a transcription of a minus strand RNA molecule from the expression cassette. Nucleic acid construct of the invention is useful for screening anti-viral drugs and determining drug resistance of an RNA virus. The present sequence is Hepatitis C virus (HCV) isolate HCV-S1 full-length polypeptide.

Sequence 3010 AA;

Query Match 96.3%; Score 1476; DB 5; Length 3010;

Best Local Similarity 95.2%; Pred. No. 8.9e-134; Indels 0; Gaps 0;

Matches 278; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

1 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 60

915 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 974

61 IFSMERKIIITWADTAACGDIISGLPVSARRREIILGPADNFBEGGKRLAPITAYSQ 120

975 VFSMEREKIIITWADTAACGDIISGLPVSARRREIILGPADNFBEGGKRLAPITAYSQ 1034

121 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAAGPK 180

1035 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAAGPK 1094

181 GPITOMTNNVDQIVGQAPPGARSMPTCTCGSSDLYLTRHADVIIVRRRGSRGSLLS 240

1095 GPITOMTNNVDQIVGQAPPGARSMPTCTCGSSDLYLTRHADVIIVRRRGSRGSLLS 1154

241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVFPVESMETTMR 292

1155 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVFPVESMETTMR 1206

RESULT 14

ABG30601 ID ABG30601 standard; protein; 2201 AA.

AC ABG30601;

21-OCT-2002 (first entry)

Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.

Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;

cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; muten.

Hepatitis C virus.

OS Synthetic.

Key Location/Qualifiers

Misc-difference 882

FT /label= Arg, Lys

Misc-difference 2183

FT /note= "wild type Met substituted by Thr"

MO200252015-A2.

04-JUL-2002.

20-DEC-2001; 2001WO-CA001843.

22-DEC-2000; 2000US-0257857P.

(BOEH) BOEHRINGER INGELHEIM CANADA LTD.

Kukolj G, Pause A;

WPI: 2002-575382/61.

Claim 3; Page: 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polypeptide region coding for a HCV polypeptide; and a 3'-NTR region. The self-replicating hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon Apk12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in the claims of the invention

Sequence 2201 AA;

Query Match 96.1%; Score 1473; DB 5; Length 2201;

Best Local Similarity 94.9%; Pred. No. 1.1e-133; Indels 0; Gaps 0;

Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

1 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 60

106 AAGIIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 165

61 IFSMERKIIITWADTAACGDIISGLPVSARRREIILGPADNFBEGGKRLAPITAYSQ 120

166 VFSMEREKIIITWADTAACGDIISGLPVSARRREIILGPADNFBEGGKRLAPITAYSQ 225

121 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAAGPK 180

226 QTRGLGCIITSLTGRKNOVEGEVQVSTATOSFLATCVNGVCMVTFHAGSKTLAAGPK 285

181 GPITOMTNNVDQIVGQAPPGARSMPTCTCGSSDLYLTRHADVIIVRRRGSRGSLLS 240

286 GPITOMTNNVDQIVGQAPPGARSMPTCTCGSSDLYLTRHADVIIVRRRGSRGSLLS 345

241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVFPVESMETTMR 292

346 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKADVFPVESMETTMR 397

